SAPIENZA
UNIVERSITÀ DI ROMA

More than words.
The structure of language as an ingredient of thought

Facoltà di Lettere e Filosofia
Dipartimento di Scienze Documentarie, Linguistiche e Letterarie
Corso di dottorato in Teorie del Linguaggio ed Educazione Linguistica

Margherita De Luca
Matricola 1419136

Relatore
Stefano Gensini

Correlatore
Grazia Basile

Index

Introduction

Why rethinking linguistic relativity (again)? Linguistic relativity is now a well-established family of approaches that has provided and still provides solid evidence of the influence of language on cognition. Nonetheless, the debate on linguistic relativity has always gone back and forth between feelings of committed acceptance and strong rejection; data speaking for linguistic relativity has been at times positively welcome, and at times refused as methodologically unstable or counterstriked on the basis of theoretical oppositions. We started to cultivate the idea that there might indeed be something in the assumptions made by relativist approaches that make them hard to integrate with broader (or different) perspectives in the cognitive sciences.

Linguistic relativity involves two key dimensions: cognition and language. If a broader framework in the study of cognition for understanding linguistic relativity is already under development in many research paradigms, the study of language – and the insights that may come from linguistic approaches - seems to still be marginalized. The definition of language and meaning was is the focus of our attention in this work. The general insight is to show that a different definition of meaning (compared to the one currently implicitly in use in relativistic studies) might help us to shed light onto the nature of the relationship between language and thought.

Allow us a personal note. Last year I had the chance to be part (as a visiting scholar) of one of the top laboratories engaged in the study of the relationships between language and cognition from a relativistic perspective. First week in my stay I presented to the lab and spent over an hour delineating the (at that time) current stage of my research: the presentation was mostly oriented towards highlighting that a Saussurian-relational view of linguistic meaning and the attention for the structure of language - for language as a system of interconnected relationships that contribute in establishing meaning - were almost completely missing in the study of linguistic relativity. The overall message I wanted to convey was, again, that the integration of such a relational perspective with research methods in psycholinguistics and the cognitive sciences would lead to a better understanding of the role of language in cognition.

I faced two major issues that day. The first one was a language issue: the problem was not as much my English, but rather the way in which I was using specific terms, like *meaning* and

semantics, which *meant* for me (coming from a continental – mostly Saussurian background) something very different from what they *meant* for them (who mostly focused on meaning as an inherently cognitive phenomenon, in line with most approaches in the cognitive sciences). The second issue found me a lot less prepared and with no substantial answer. At the end of my presentation, it was the time for questions. The professor at the head of the group, who was hosting me in the lab, thanked me, smiled at me, and then made a comment that sounded more or less like this: "even supposing that you are right, and hence that linguistic relativity approaches should use a different working definition of meaning, what I am supposed to do with that?" The criticism was evident and striking: I had no idea, not even a general plan, on *how* to integrate a relational, systemic view of meaning in relativistic research paradigms. This, in turn, meant that, if I wanted to bring this perspective on meaning into the study of linguistic relativity, I had to bridge the gap between a purely theoretical account and the needs of psycholinguistic methodologies, i.e. the relational system of language had to be operationalized in order to become a suitable ground for the formulation of testable hypothesis and, essentially, to be manipulated as variable in experimental setups. This work should be understood as a first - definitely open-ended - attempt to answer this question.

In this brief introduction we will try to summarize how the answer to this question brought us substantially far from (strictly speaking) linguistic relativity. On the one side, this attempt to integrate a Saussurian perspective with relativistic studies led us to a broad reformulation of the claims underlying the relativistic inquiry and, on the other side, it invoked the need for a much broader discussion of the relationship between language and thought. As we will try to show, what we end up with at the end of our discussion is not a neo-Whorfian account of the influence of language on thought, but rather a (to some extent) new possible direction of research in understanding the problematic interface between language and thought.

In a nutshell, this work is mostly about defining the key components of the debate in line with current approaches in linguistics and in the cognitive sciences. We will try to show how a renewed definition of the terms leads to a considerably new perspective on the kind of relationships and influence that may hold between language and the rest of our cognition. Here, we will just briefly outline some of the main lines of research taken over in this work, and we will postpone a more detailed presentation of the specific content of each section to the end of next chapter as, by then, we should have a better understanding of *why* we engaged in this work and *where* it should

lead us to. As for now, we will try to answer a very simply - but at the same time crucial – question. What is linguistic relativity actually about?

If we oversimplify a lot and try to narrow the issue to its bone, linguistic relativity is mostly about language diversities. After the Chomskian revolution in linguistics and the rise of cognitive sciences, it has been generally taken for granted in cognitive and psycholinguistics approaches that languages, at least in the deep structures beneath their surfaces, show substantial universality and uniformity. Linguistic relativity represented a theoretical and methodological alternative to the Chomskian universalism. The mere fact that historical-natural languages are diverse and that these diversities are pervasive posed (and still poses) fundamental questions to the universality of both language and, more broadly, of human cognition.

And this seems to be true not just in the realm of scientific studies. Language diversities have always been a topic in the science fiction, both when acknowledged and exploited for many different reasons or when – almost programmatically – ignored. For example, in *Rick and Morty* – an American adult animated science fiction sitcom - a brilliant scientist and his not-very-apt-to-the-task nephew constantly travel across the universe and across time. No matter if they end up in Blitz and Chitz, an intergalactic arcade center, or if they watch an intergalactic cable with many funky episodes, every single individual in every single one of those universes appears to speak perfect English. No accents, no variations in word choice, no different languages. Conversely, language diversities seem to be a big issue for galactic hitchhikers in Douglas Adams's *The Hitchhiker's Guide to the Galaxy*, where a specific device – the *Babel fish*, a small, bright yellow fish - when placed into someone's ears allows them to hear any language translated into their first language. In the movie *Arrival,* based on Chiang's short story *Story of your life,* the aliens that land on Earth talk – and think - about time in a non-casual way; in learning the alien's language, the main human character in the movie starts to view and experience reality – and especially time – in a very different way, consonant with the alien's linguistic categories.

These three examples represent three different (taken to the extremes) ways in which the scientific community has been dealing with language diversities over the last decades. In the first case, diversities are simply not an issue and the study of the English language might as well serve as a universal gate to the study of language (and thought) in general. In the second case, diversities definitely exist, they have to be acknowledged, and there are ways in which we can overcome them making (out of metaphor) language an instrument, and not a prison, of thought. In the third case –

the deterministic case - language diversities are so foundational to our constitution of reality that they shape it so deeply we cannot escape it. Much of the research we will present lies somewhere between these three vertices. The reason why we are highlighting this here is that (at least historically) the perspective assumed on language diversities generally goes hand-in-hand with specific assumptions (or thesis) about the nature of the relationship between language and thought. The more language is taken to simply mirror cognition (and specifically concepts), the more language diversities are downplayed. Conversely, the more we accept the presence of outstanding differences among languages, the more we will lean towards an active cognitive role for language.

But even more interesting is that, if we come out of the science fiction territory, language diversity, the possibility of translating across languages, and the influence of such diversities of thoughts, are not just topics for late night discussions over a bottle of good wine. It is now widely acknowledged that language diversities constitute the very reality in which we are embedded.

What does it take – in theoretical as well as methodological terms – to understand language diversities and to scout the influence of language on thought? To what extent and in what way do languages differ? If the language we speak indeed influences thought, how should we understand this influence? In the first introductory chapter we will reconsider linguistic relativity in the light of these question and we will try to show how the main underlying issue is the very possibility of distinguishing between language and thought. The fact that the two cannot be considered as being identical comes right from the observation that languages differ (we will develop this argument in chapter 1). Here is where the Saussurian perspective on meaning comes handy. Languages do not simply label a pre-organized world or a pre-given set of concepts. Languages differ in the way they carve up both the external as well as our conceptual reality. The way in which languages are organized - the structural and relational aspects of meaning - gives us the starting point to rethink some of the main issues in linguistic relativity and, more specifically, to rethink how language relates to thought. The first and second chapters of this work will be entirely devoted to this. We will first focus on the history and on more recent developments in linguistic relativity, trying to show *why* we need a different account of meaning in order to ground – weaker or stronger – relativistic claims. We will see how language diversities will play a central role in the structure of the argument, and how the level of the structure of language will offer interesting insights onto the relationship between language and thought.

At the end of the first two chapters, we will be able to draw the following conclusions: languages are different; languages organize our experience of the world in language-specific ways; linguistic meaning is differential and relational; our discussion of the language-thought interface should start from these fundamental insights. So far (almost) so good.

Moving forwards, the next step will be to identify how and where language and thought differ, in order to later discuss how the two may influence each other. The discussion I will propose leads to the possibility of identifying a "language-specific" level of linguistic meaning - intertwined but not reducible to other cognitive aspects of meaning formation – in the relational aspects of meaning as identifiable by means of distributional semantic approaches. This move allows to drag – so to speak – the relational systemic view of meaning into the lab, i.e. to make the notion of language as a system operationalizable. Chapter 5 will focus on how distributional semantics has its roots in the Saussurian soil. Chapter 6 will focus on how insights from distributional semantics allow us to reframe the language and thought interface and the influence on the first of the latter in a way that is coherent with current advances in linguistic and psycholinguistic approaches and in computational techniques. We will try to show how a distributional account of the language system is in line with the intuition that language is one ingredient in the formation of our mental content and, hence, that in such way *language plays a role in shaping our thoughts*.

One note is here in order. Operationalizing the whole notion of meaning and of language as a system as Saussure - or as post-Saussurian thinkers presented - it is clearly impossible. The range of notions, intuition, reflections, proposals, and problems related to a structural/structuralist systemic view of meaning is so broad that probably no single theory can even get close to encompass them all. Even less if the aim is to make the systemic view of meaning suitable for manipulation in empirical studies. Hence, the first step was to narrow down the area of interest and to select (or at least highlight) specific instances within the broader picture of the Saussurian framework. Here, we will mostly focus on the relational view of meaning, i.e. on how the meaning of a linguistic units depends (at least partly) on its interrelation with other units in the system. We will see how recent advances in computational linguistics, namely in distributional semantics, allow us to do precisely that. We are well aware that this does not settle for good the issues we are about to raise.

It should also be noted upfront that we will progressively lose the focus on language diversities – that nonetheless remain the starting point of the argument. The first two chapters should shed light onto why this is the case. Rethinking the notion of meaning underlying linguistic relativity approaches will lead us to a profound reconsideration of the principles underlying the relativistic claim. What we will mostly focus on is the possibility of disentangling between language and thought and on the consequences of identifying a "language-specific" level of meaning in the relational aspects of language.

Chapter 1

Linguistic relativit(ies) and the study of meaning

At this year's *Annual Meeting of the Cognitive Science Society (41st)*, one of the five renowned Glushko Dissertation Prizes for outstanding theses in cognitive sciences was awarded to a dissertation entitled: *Language, meaning, and visual perception: Event-related potentials reveal top-down influences on early visual processing*.[1] Needless to say, Maier's thesis talks about linguistic relativity (about if and how the language we speak shapes the way we think and perceive); so much so that at the end of his presentation to the audience, the author jokingly stated that the cognitive sciences should not be afraid of linguistic relativity anymore. Of course, this dissertation does not simply restate the Whorfian or strictly neo-Whorfian claim of the relativity principle, but rather something more specific and more general at the same time. Neuroscientific evidence reveals that early visual processes (i.e. processes involved in the so-called lower level cognition) are open to influences from higher-level cognitive abilities (hence, we observe top-down effects on, for example, perception). Maier investigated one specific aspect of higher-level cognition: language, and, more specifically, the categories provided by the language we speak. The main domain tested by the author is color perception and categorization. Within this framework, linguistic relativity is reframed and finds a new theoretical and empirical ground: linguistic relativity can be seen as a consequence of the general organization of cognition. Language influences the way we think and perceive because the way in which we perceive the world is influenced by the categories and structures of higher-level cognitive abilities, one of which is language.

Maier's award tells us some very important facts about (some of) the current interests in cognitive sciences. In the first place, the fact that one of the most renowned prizes for young researches was awarded to a dissertation concerned with linguistic relativity is impressive by itself: linguistic relativity is still (or most accurately, is again) a central topic in research in the cognitive

[1] For a description of the prize and information about 2019's and past year's winners see https://cognitivesciencesociety.org/glushko-dissertation-prize/

sciences. This is already remarkable, if we consider how much skepticism the study of linguistic relativity has received throughout its (brief) history in this field. In the second place, this award tells us that if we want to think of linguistic relativity in cognitive terms we might want to look for new foundations: linguistic relativity effects can be explained within a broader framework for understanding our cognitive architecture and functioning and can be supported by neuroscientific evidence. In the third place, it tells us (together with an ensemble of approaches we will discuss shortly) that, if cognition is the central focus of research, much less effort is devoted to the study of language.

The linguistic relativity principle (or hypothesis) inherently involves two main key dimensions: cognition and language. As we will see, a third central aspect of the claim is how the two relate to the external world. The idea that language influences thought can be understood as a specific instance of the study of how the two, language and thought, relate to each other. It seems thus intuitive that a preliminary understanding of the nature of the two parties involved is a fundamental prerequisite if we aim at studying the relationships they entail. In this sense, having a theory of cognition and a theory of language is fundamental to develop a theory of language in cognition. And even if studies in linguistic relativity do not develop a full-fledged theory of cognition, language, and meaning, they necessarily rest on some (often implicit) assumptions about the nature and functioning of the dimensions involved. As part of a broader research field that aims at investigating and shedding light onto our cognitive abilities, researchers interested in the linguistic relativity idea within the cognitive sciences have mostly been concerned with a theory of cognition, its structure and its functioning, and much less with the study of language. However, their works rest on often implicit assumption of what language is, how it works, and how language conveys and structures meaning. This observation is by no means meant as a critique. In very general terms, research in the cognitive sciences aims at unveiling how cognition works and, from a cognitive perspective, language is essentially another aspect of our cognitive abilities that should and needs to be investigated within psychological terms. The peculiar aspect of relativistic approaches, if compared with other cognitive accounts of language, is that they view language as playing an active cognitive function, whereas most approaches in the cognitive sciences tend to reduce the role of language to its communicative function. According to these latter approaches, meaning is conceptual, and language essentially maps onto the structure of concepts. In contrast,

the linguistic relativity framework sees language as playing a key role in shaping – in giving rise to - our conceptual knowledge.

Hence, on what notion and definition of language and meaning are relativistic studies based? Let us reveal in advance that this question has no easy answer; studies in linguistic relativity often lack a precise definition of language or meaning. From our point of view, this lack jeopardizes the relativistic enterprise, which, in turn, would benefit from an in-depth discussion of how language, together with the mind, makes meaning. It is clear that talking about language and meaning inevitably leads to talk about cognition; language and meaning are inherently cognitive, in the sense that they are an essential part of our cognitive abilities and necessarily depend upon them. How, then, is our argument not circular? In this thesis we want to evaluate the possibility that a linguistically informed study of language and meaning may help us shedding light onto important aspects of language in cognition and onto the relationship between language and thought, in the same way in which neurological and behavioral data can tell us about the nature and functioning of this relationship. In very general terms, we are interested in evaluating the role of language in cognition from a precise standpoint about the nature and organization of language: meaning in language is relational, and hence is (at least in part) given by the web of interconnected relationships between the units of language. Our proposal is that the investigation of language from the point of view of linguistics (namely, Saussurian linguistics) can bring important insights into the study of language in cognition and may lead to a reformulation of linguistic relativity grounded on specific, openly theorized, ideas about the nature and structure of meaning.

Before turning to a brief introduction to our proposal, of which the linguistic aspects will be better addressed in chapter 2, we want to introduce the notion of linguistic relativity and to outline its history and current state of the art, mostly focusing on the notion of meaning that underlies traditional as well as contemporary approaches.

One further consideration is here in order. We will start by discussing linguistic relativity, but we will hopefully be able to progressively show how the relativistic claim is part of a much bigger and entangled question about the relationship between language and thought, about the nature of linguistic meaning, and about its role in shaping our semantic and conceptual representations. In the next chapters, we will discuss each of these terms, trying to point out our theoretical as well as methodological route. Starting from linguistic relativity allows us to frame the debate that follows, outlining to what respects studies in linguistic relativity have been

grounding their research in an oversimplified notion of language and meaning. Furthermore, it seems rather clear that, since the original formulation of the relativistic principle in the Whorfian and in the cognitivist frameworks, much has changed in what we know about both language and cognition. We will show how the relativistic debate has been following this more general development in the study of language and cognition, reshaping its theoretical claims as well as its methodological framework. Finally, we will try to draw some general conclusions about what linguistic relativity is at and what might be, according to the view of language we will endorse, a profitable line of future research.

1.1. Linguistic relativity in the history and in the present

1.1.1. From anthropology to cognitive sciences

In its most famous formulation, linguistic relativity is the idea that the way we speak shapes the way we think (and to some extent, the way in which we perceive the external and even our internal world). The main reason that led researchers to endorse one of the various possible declinations of the linguistic relativity idea is that language diversities pose fundamental questions to the supposedly straightforward and neat universality of cognition and language. Language diversities undermine the view of language as a mere means of communicating thoughts, leading the way towards the study of a possible cognitive function of language.

Before it became a cognitive hypothesis in search for empirical testing, linguistic relativity was a theoretical principle[2] based on strong assumptions about the nature of language. Linguistic relativity was initially founded within the anthropological-linguistic tradition that went from Boas, to Sapir, and, finally, to the full development of the linguistic relativity principles in Whorf. Nonetheless, its main theoretical foundations have deeper roots in the history of philosophical and linguistic ideas and can be traced back to, at least, the thought of Wilhelm von Humboldt as theorist

[2] Viewing linguistic relativity as a principle and not as a hypothesis was one of the main arguments proposed in the historical re-evaluation of the theory after the Chomskian area, which essentially denigrated the theory and put it on hold. For instance, Hill and Mannheim "maintain that "linguistic relativity" as proposed by Boas, Sapir, and Whorf is not a hypothesis in the traditional sense, but an axiom, a part of the initial epistemology and methodology of the linguistic anthropologist" (1992, p. 383). According to this perspective, linguistic relativity is not a hypothesis that needs to be tested, but rather a starting point from which to draw experimental hypothesis and lines of inquiry. For other attempts to reframe and interpret the original relativistic claim cf. Black (1959), Kay and Kempton (1984), Gumperz and Levinson (1996a), Hill (1988), Lee (1979).

of the *Verschiedenheit*.[3] The Humboldtian roots of linguistic relativity are often misinterpreted or forgotten, but it is indeed by means of their penetration and perpetration in the thought of Franz Boas (1985-1942) that they became one of the constitutive pillars of the subsequent notion of linguistic relativity. From its very foundation, cultural anthropology (which was started in the second half of the Seventeenth century by Edward Burnett Tylor) has offered the most fertile ground for the rise of the relativistic view of language. The leading figure in the discipline, the German anthropologist Franz Boas, which emigrated in the United States in 1887 and was involved in a long series of journeys and fieldworks as well as academic positions, up to becoming full professor at Columbia University, dedicated groundbreaking works to those American-Indian languages that had already encountered the interest of Humboldt for their polysynthetic and incorporating structure, that so much differed from the well-known and usually studied Indo-European languages. In the work of Boas, the Humboldtian principles were filtered through the lens of the rising discipline of cultural anthropology, which provided methods and theoretical arguments for the observation of language diversities and grounded the study of languages as historical and cultural phenomena. Crossing the ocean with a theoretical background highly entrenched in Herder's and Humboldt's philosophy of language, Boas was the first who looked at the so-called "primitive" languages not in comparison to the gold-standards of Indo-European languages, but rather within their own particularities. Boas clearly highlighted that different languages diverge in the way they habitually classify experiences and segment the natural and cultural world,[4] although without explicit arguing that such differences may in turn be the source of differences in the ways people think.[5]

[3] In this brief introduction on the notion and core ideas of linguistic relativity, we will not offer a full account or discussion of its Humboldtian roots. For a review cf. for example Deutscher (2010) and De Luca and Gensini (2019).
[4] The most famous example is that of the terms for *snow* in Eskimo. For a history of the debate and some of the major critics cf. Pullum (1989) and Pinker (1994). In underlying the triviality of the snow example and the cultural dependence of such observation, critics to the snow example undermine the focal point, being the connection between lexicalization and categorization: "observations of novel lexical patterning or elaboration can serve as the raw material for hypotheses and can inform theories of human cognition in the same way as patterns of associations and dissociations observed through neurological case studies can inform theories of human cognition. It was this connection between labeling and categorization that formed the crux of the original Eskimo snow example by Boas. Using a word involves a choice to select certain aspects of the experience. Being part of a community that uses a particular word requires the learner to perform the necessary acts of categorization to be able to use the word properly" (Lupyan, 2012a, pp. 263–264).
[5] For a discussion of Boas's ideas and their role in shaping the future notion of linguistic relativity cf. Koerner (1992) and Lucy (1992).

In the works of Edward Sapir and Benjamin Lee Whorf, the original Boasian interests and attention for the study of language diversities fed the idea that language is not just a mirror of the external reality, but rather that different languages "read" what is supposed to be the same reality in often very different ways, according to their own categorial schemes. In this sense, language is not just a *Weltbild*, a picture of the world, but rather it carries a *Weltansicht*, a view of the world. This idea is already present in Sapir's first works and completely unfolds in *Language* (1921) and some subsequent essays. In 1924, Sapir explicitly refers to "a kind of relativity" in the way in which we engage with reality, stemming from the unaware adoption of "fixed habits of speech" (1924, p. 155) embedded in our native language system. It is the systematic nature of language, i.e. the fact that language is co-extensive with the culture of a given society, that shows us how language wraps our individual as well as collective experience. In *The Status of Linguistics as a Science* (1929), Sapir stated that the language we speak is a filter between the individual and the external world; in fact, the latter cannot be thought of as objectively pre-given to us, but rather as inherently constituted by means of the linguistic habits of a given speaking community: "no two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached" (1929, p. 209). Sapir did not indent to deny the underlying principle that equates all human being beyond their cultural and linguistic differences; the principle (Humboldtian first and Boasian then) of the universality of our experience and knowledge is the other facet of the process of specification that historically characterizes both. However, whereas Boas did not go beyond acknowledging that culture and language reciprocally mirror each other, Sapir hypothesized an active role of language in the sense that, once it is established as an organized whole (and it is precisely in this way that the individual speakers acquire it), language predisposes its speakers towards certain interpretative choices instead of others, hence becoming a "symbolic guide to culture" (1929, p. XXX). Therefore, as Sapir already noticed in *Language*, if culture can be viewed as *what* a given society thinks and does, language is the *how* that knowledge is organized.

The idea of language as an organized totality that does not simply mirror the external world, but rather provides a guide to it, foreshadows the way in which Whorf gave form to the linguistic relativity principle[6]:

> we dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds – and this means largely by the linguistic systems of our minds (Whorf, 1956, p. 213).

As his predecessors and mentors, Whorf was a careful observer of differences in languages and recognized the pervasiveness of their structural differences. The linguistic relativity principle is first and foremost grounded in the centrality of the classificatory nature of language in regard to experience and on the internal systematicity of languages: each and every linguistic distinction works in accordance with other distinctions, giving rise to *fashions of speaking,* in the formation of a coordinated and coherent whole. As a consequence, the differences between languages can be understood and characterized in terms of broad schemes of differences that cut through the system. In the Whorfian approach, the key role in the definition of relativity is played by the structural aspects of language and by the extensive observation that different languages exhibit different structures. According to Whorf, the structure of language depends on the interrelation of *overt* and *covert categories*, which diverge in the level of formal marking (the first is immediately evident in the structural organization of a language, the latter needs to be reconstructed from the organization of the system), and together give rise to a complex interrelation that profoundly shapes the way in which we conceptualize (and even perceive) reality. The famous comparison between the conceptualization of the domain of time in English and Hopi (of how the two languages temporally organize experience) remains one of the best examples of a relativistic study. The comparison is not restricted to the evaluation of the differences in the words that can be taken to denote time, but it spreads to the analysis of grammatical patterns and, more broadly, to those *fashions of speaking* that underlie the *metaphysics* (the view of the world) of a given linguistic

[6] Since the linguistic relativity principle was first articulated, or at least first quasi-cohesively articulated, in the work of Whorf and his teacher Edward Sapir, it is often referred to as the "Sapir-Whorf hypothesis". This label was first introduced by Carroll in his introduction to the 1956 edition of Whorf's writings. However, it should be mentioned that, even if Sapir's reflections indeed opened up the way to the formulation of the relativistic principle, his thought cannot be taken as overlapping with Whorf's proposal on the influence of language and thought. For a review of the main points of contact and differences between the two authors cf. Lucy (1992).

community. Hence, each language offers its own way of structuring and organizing the external world and the individual speaker, by means of being embedded in a given language, acquires this linguistic way of conceptualizing the world.

Although Whorf never explicitly endorsed a strong deterministic view of the role of language in shaping cognition (we should not forget that one main tenant of Boas, Sapir, and Whorf was the underlying universality of the human nature), the Whorfian perspective is often taken as the origin of the strongest version of the relativistic claim: linguistic relativity implies the incommensurability of the different *Sprachansichten* (of the linguistically informed views of the world) and the rise of the formative role of language above and beyond the human biological limits and the autonomy of the structure of the world. This version of the linguistic relativity principle essentially became the target of the generativist *nouvelle vague,* which led to a substantial downfall of the research and interest in the relativity principle. The innate and universalist premises of the Chomskian paradigm, together with the refutation of the centrality of semantics in the study of language (e.g. Chomsky, 1957), seemed to subtract any credibility to the relativistic claim. Generative approaches substantially devalued the role of the (cultural) acquired components of language, which is reduced to the parametric modulation of an innate and universal base of combinatorial rules. Therefore, semantics was eradicated from the core of language and relegated to a superficial "interpretative" function. Moreover, language was denied any meaningful formative role in cognition: the role of language is simply to label thoughts (however they might be defined) and to allow to communicate them across speakers. The generative paradigm straightforwardly endorsed the idea of a priority of the cognitive over the linguistic aspects: meaning is *internal* to the mind and it depends on the manipulation and processing of symbols. Hence, the semantic structure of language could be viewed as a direct window on the structure of thought (for a discussion cf. Goldin-Meadow, 2003). The critique to the relativistic claim was strikingly powerful: languages do not differ one from the other in any substantial sense given that their core structure is innate and universal, and semantics is a mere surface effect; language is a separate module in the mind, which is not implicated in the formation of our mental content, but rather simply reflects it. Even though the generative paradigm targets a version of the relativity principle that no one (not even Whorf) probably ever argued for, i.e. the strong deterministic claim, the rise of the generative paradigm and the widespread favor it received at the very formation of

the cognitive sciences were enough to dismantle the relativistic claim and charge it as either being nonsensical or too trivial to even be addressed.

This parable of ascent and decline of the relativist claim is exemplified by the case of the color domain, which has been the test-bed for both universalist and relativistic claims over the centuries.[7] An anti-relativistic study that remained classic in the literature on the perception and categorization of colors is Berlin and Kay's *Basic Color Terms* (1969), in which the authors showed that the relativistic claims had to be substantially reshaped in the light of the uniformity of the color perception across cultures, which essentially determines the focal color around which color categories are constructed and to which they are indeed bounded. The way in which we perceive and categorize the color domain is a consequence of our biological structure and not so much of the language we speak ([Rosch] Heider, 1972). Hence, what was under attack was the very idea of semantic variability and, with it, one of the fundamental assumptions of linguistic relativity. However, despite the tenacity with which linguistic relativity was rejected[8] and the inherent limitations of the Whorfian formulation, often largely obscure and underspecified, linguistic relativity underwent a second consistent wave of interest starting from the early 90s and stemming from within the cognitive science, i.e. from approaches within the same paradigm that seemed to have proclaimed its defeat just a couple of decades before.

The author that mostly favored and contributed to the revitalization of the study of linguistic relativity was the American linguist and psychologist John A. Lucy, who in 1992 published the groundbreaking volume *Diversity and Thought: A reformulation of the Linguistic Relativity Hypothesis*, followed in 1996 by *Grammatical Categories and Cognition*; the first volume offered a thorough re-evaluation of the history of linguistic relativity from both a theoretical and methodological standpoints; in the second volume, the author presented the empirical work carried out on Yucatec Mayan starting from the 70s. Lucy's work probed to be fundamental inasmuch as it explicitly posed the foundations for a psychological inquiry into linguistic relativity. Lucy showed how already in the 50s, and hence in conjunction with the rise of the Chomskian paradigm, attempts were made within the cognitive sciences to merge the Whorfian claims with the theoretical and methodological requests of the new-born sciences of

[7] For an extensive reconstruction of the relativistic versus universalist debate on the color domain cf., among others, Deutscher (2010).

[8] See, for instance, the caricatural account that Pinker (1994) gives of the linguistic relativity and the response in Casasanto (2008).

cognition. Furthermore, he demonstrated how a cognitive approach to relativity was already implicit in the Whorfian formulation, which, if properly understood and stripped from the untenable deterministic claims, offers substantial and testable hypothesis on the role of language in the way in which we engage with the world.

How may this integration of linguistic relativity and cognitive sciences work? Lucy (2016)[9] sees in the rise of psycholinguistics, and especially in the works of Eric Lenneberg and collaborators (R. W. Brown & Lenneberg, 1954; Eric H. Lenneberg, 1953), the second foundational moment of linguistic relativity that essentially reshaped the relativistic claim and turned the principle into a testable hypothesis. The problem with the Whorfian claim, as highlighted by these authors, is both theoretical and methodological: given that linguistic relativity hypothesizes the way and the direction of influence of language on thought, then this influence should be observable and observed in the study of cognitive processes. The general critique was that Whorf theorized the influence of language on thought without actually ever going beyond linguistic data. In contrast, according to Lenneberg and his collaborators, differences in languages (in linguistic domain such as color, time, and space as well as in language structures such as how different languages assign gender to nouns or how they treat numeral markings) would give rise to differences in non-linguistic behavior, such as perception, attention, long-term or short-term memory and so on.[10] Hence, the hallmark of the new approach was that of looking for the so-called Whorfian effects on performances of individual speakers by evaluating, by means of experimental psychological evidence, what aspects of our linguistic representations interfere with what mechanisms of non-linguistic cognition, giving rise to linguistically-shaped non-linguistic behavior. This second foundational moment can be seen as the "cognitivization" of the Whorfianism and it delineated the field of all the subsequent work in linguistic relativity, focusing the experimental enterprise not so much on linguistic differences and different worldviews, but rather on pinpointing what cognitive processes and what variation of those processes may be ascribed to linguistic manipulations established in experimental setups. It seems clear that

[9] See Lucy (2016) for the explicit distinction between first and second foundational moments, but the idea and the historical reconstruction is already fully developed in his masterpiece (Lucy, 1992).

[10] For instance, Brown and Lenneberg (1954) investigated the linguistic codability of the domain of colors: starting from an objectively measurable referential domain of experience (e.g. color) on which both language (i.e. language established names for colors) and cognitive mechanisms (e.g. memory and perceptual recognition) operate, Brown and Lenneberg showed a positive correlation between linguistic codability of colors and participants ability of recognizing and remembering them.

something substantially changed with the foundational aspects of the relativity enterprise: namely, the definition of language. In its psycholinguistic reformulation, the relativistic view loses that primary interest for linguistic diversities and for the systematic and structural aspects of language, which were indeed prevalent in the Whorfian original claim.

While maintaining the need to tackle onto non-linguistic processes and behaviors to guarantee the individuation of the mechanisms by which language operates its influence on thought, Lucy regains two core aspects of the Whorfian claim: on the one side, the necessity of a contrastive approach[11] and, on the other side, the attention for the structural (mostly grammatical) aspects of the examined languages. The empirical method constructed on the basis of these assumptions addresses and evaluates the cognitive implications of the structural differences observable in languages:

> [s]uch research should be comparative in that it should deal with two or more languages. It should deal with a significant language variable such as one or more central grammatical categories rather than a relatively minor vocabulary set. It should assess the cognitive performance of individual speakers aside from explicitly verbal contexts and try to establish that any cognitive patterns that are detected also characterize everyday behavior outside of the assessment situation (Lucy, 1996b, p. 48).

The majority of the subsequent studies has readily taken over the need of a comparative approach as the basis of any relativistic claim, triggering a flourishing multiplicity of cross-linguistic analyses. A somewhat different fate awaited the attention posed by Lucy to the structural and grammatical aspects of language. The majority of contemporary approaches focuses on the analysis of very specific and narrow lexical domains and on their influence on various specific cognitive abilities. To this regard, Lucy (2005) proposes a classification of contemporary approaches into two main categories: *domain-oriented* approaches focus on the analysis of specific lexical domains; *structured-oriented* approaches are oriented towards the analysis of grammatical structures and their influence on cognition. This latter category comprises, for example, Boroditsky and colleagues' (cf. Boroditsky, Schmidt, & Phillips, 2003; Phillips & Boroditsky, 2003) studies on the role of grammatical gender and of how differences in the way in which it is expressed or

[11] Lenneberg and collaborators programmatically focused only on English on the basis that what was under scrutiny was the very possibility of the influence of language (any language) on non-linguistic cognitive processes.

not expressed in language may influence our way of conceptualizing and representing objects[12], and Lucy's (1996a) own studies on numerical marking in English and Yucatec Maya and the consequent differences in cognitive processes. Far richer is the pool of *domain-oriented* studies that now cover a wide variety of domains from classical themes such as time, space, and colors, to more recently addressed domains such as body parts and numerals.

Even in domains that are considered as basic and foundational in our experience of the world, such as space and spatial organization, we are faced with a considerable variability in how languages encode it. For instance, Levinson and his group at the Max Plank Institute for Psycholinguistics of Nijmegen showed that the system of spatial references languages provide their speakers with (intrinsic, relative, absolute, or a combination of the three) leads to differences in cognition consistent with the linguistic organization (Levinson, 1996, 1997, 2003; Pederson et al., 1998).[13] In the domain of time and of the spatial metaphors employed to represent it, Boroditsky (2001) showed a correlation between schemes of lexicalization and non-linguistic behaviors starting from differences observable between metaphors for space in English, which orients time horizontally, and Chinese Mandarin, which orients time vertically. English speakers were faster at verifying or falsifying temporal sentences with spatial metaphors if primed with language-consistent (hence, horizontal) spatial stimuli; similar results were found for Chinese Mandarin speakers in verifying and falsifying vertical metaphors.[14] Lastly, in the domain of color categorization and perception, a recent and emblematic study is that of the *Russian blue*. In contrast to English, Russian (as well as Italian) separately lexicalize dark blue (*siniyi*) and light blue (*goluboy*) and this difference correlates with differences in the ability of speakers in distinguishing between the two areas of the chromatic space (Winawer et al., 2007). This study shows that the categorical distinctions operated by language may have an influence on our abilities of perceptually discriminating between the two. Finally, Maier (the winner of the Glushko prize we mentioned before) and collaborators (2018) showed the influence of linguistic categories on early visual perception at a neurological level by pinpointing it in the first milliseconds of our perceptual activity.

[12] Various subsequent studies have revealed that the interpretation of the results provided in the above-mentioned studies should be variously softened and that the effect of language on how we cognitively represent objects may indeed be just of a superficial kind (for a review cf. Lalumera, 2013, pp. 77–84).
[13] For a critical account of these studies and of the results cf. Li and Gleitman (2002).
[14] For a moderate claim on the role of language in shaping time conceptualization cf. January and Kako (2007).

The list of studies that speak for (as well as against) this idea of linguistic relativity is now extremely long, and a full review of the history and present state of research and of over the years accumulated data could easily take up the rest of our work. But this is not our goal.[15] In this section, we have briefly outlined the history of the linguistic relativity principle from its (North-American) birth to some of its most recent developments within the cognitive sciences. The way in which the relativistic issue was framed within this paradigm is notably expressed by Everett: "do patterns of thought vary in accordance with one's native language? Put differently, does there exist a sort of linguistic relativity, such that some aspect(s) of a person's cognition depends on, or is relative in accordance with, the language employed by that person?" (2013, p. 1). The neo-Whorfian answer to this question can be summarized as follows: linguistic representations interfere with non-linguistic representations in giving rise to linguistically-shaped non-linguistic behavior; languages exhibit substantial semantic variation; different linguistic representations lead to different linguistically-shaped non-linguistic behaviors (cf. Gumperz & Levinson, 1996a, p. 24). Let us move one step further before turning to some general conclusion, and briefly present other clusters of approaches within the cognitive sciences that appeal to linguistic relativity but take slightly different theoretical and experimental routes from the one presented above.

1.1.2. From relativity to relativit*ies*

Researches in linguistic relativity are far from being a unitary approach, even if considered just within the restricted framework of the cognitive sciences. In fact, not only one can isolate a specific linguistic domain and/or cognitive process and investigate the possible effects of language, multiplying the number of possible studies, but there are also different types of linguistic effects on cognition that can be individuated, and very different mechanisms can be thought of being involved in giving rise to such effects. Hence, instead of addressing the issue of linguistic relativity we should more appropriately offer a discussion of how the various types of relativistic approaches frame the general question of the influence of language on thought.

We observe this multiplicity of effects and mechanisms already in Whorf's work. One type of evidence presented by Whorf relates to the way in which languages "dissect nature differently"

[15] For in-depth reviews of the state of the art of researches in linguistic relativity cf., among others, Everett (2013) and Lalumera (2013).

(Whorf, 1956, p. 208). There is (especially now) extended evidence that different languages provide speakers with different categories or lexical items for describing the same portion of reality. This kind of linguistic differences give rise to a sort of "categorization effect" (Everett, 2013, p. 16): the way in which a particular semantic field is divided up in a given language impacts the way in which speakers conceptualize it, or even perceive it. One of the corollaries of this proposed effect is that speakers with a greater number of basic lexical items for a given semantic field construe that semantic field in more precise or discriminating ways than speakers lacking such terms. A similar effect of linguistic categorization can be observed in perception, where greater perceptual refinement is implied by more detailed lexical demarcations between portions of a given semantic domain. The other kind of relativistic effect suggested by Whorf is a sort of "analogy-based" effect (Everett, 2013, p. 17): the meaning of words could lead to non-linguistic behaviors that are more or less directly induced from the meaning of the words themselves, and these analogy-based behaviors would most probably not be observed in speakers of languages that do not share the same meaning.[16]

As we have briefly shown in the previous section, Whorf's speculations were just the very beginning of the cognitive exploration into linguistic relativity, which can now be said to comprise a 'family' of related although different proposals. Therefore, even though in the previous section we have presented linguistic relativity as one specific approach to a much broader question, we should now expand the notion of relativity and try to look at it from the many different perspectives in which it has been investigated. In recent years, there have been many attempts to offer systematic reviews of the different "types" of approaches that fall under the general umbrella of "linguistic relativity". We will briefly review some of the most well-known classifications of approaches into various macro-categories that should give us a general intake of the broad space covered by linguistic relativity nowadays. Aside from this general overview, we will focus on one specific kind of relativity that is gaining theoretical consensus and that is reaching interesting methodological achievements, namely the *language-augmented thought* proposal. We will see

[16] The classical Whorfian example of this kind of relativity is the meaning of the word "empty" and the behaviors it induces in English speakers, even when those behaviors are indeed very risky. "Thus, around a storage of what are called "gasoline drums," behavior will tend to a certain type, that is, great care will be exercised; while around a storage of what are called "empty gasoline drums," it will tend to be different – careless, with little repression of smoking or of tossing cigarette stubs about. Yet the "empty" drums are perhaps the more dangerous, since they contain explosive vapor. Physically the situation is hazardous, but the linguistic analysis according to regular analogy must employ the word 'empty,' which inevitably suggests lack of hazard" (Whorf, 1956, p. 135). We will come back to the mechanism that is taken to underlie this kind of effect in paragraph 1.2.3.

how this perspective opens to a variety of insights into the study of the role of language in cognition that, to some extent, differ from the neo-Whorfian formulation we reviewed in the previous section.

Let us begin with an outline of various possible "types" of linguistic relativity. One very general categorization of "types" of linguistic relativities is the one proposed by Lucy (e.g. 1996b), who delineates three main categories: "semiotic", "structural", and "discursive". The first relates to the question of whether language "fundamentally alters the vision of the world held by humans in contrast to other species" (1996a, p. 39). Structural and discursive relativity depend on the first but address more specific phenomena. Discursive relativity refers to the notion that variations of patterns in the usage of language may influence the thought patterns of particular speakers. Even within particular linguistic communities, usage varies across speakers for functionally oriented reason (e.g. Hymes, 1974) and this sort of variation is likely to give rise to specificities in thought patterns. Structural relativity refers to the kind of relativity we have discussed so far, i.e. the influence of language (linguistic categories or language structure) on non-linguistic thought processes. Within the broad category of "structural relativity", relativistic claims have often been organized along a continuum from very week to very strong (i.e. linguistic determinism) claims (e.g. R. Brown, 1976). An example of a weak kind of relativity is Slobin's (1996, 2003) idea of *thinking for speaking*, i.e. the idea that speakers must engage in a special kind of mental activity when using language: the effects of language on cognition would be limited to those cognitive activities that directly involve language or to the thinking that occurs immediately prior to using language (i.e. thought processes involved in speech production) (cf. Wolff & Holmes, 2011).

Although relativistic approaches can definitely be taken as endorsing weaker or stronger claims, they can hardly be set onto a weak-to-strong continuum inasmuch as they vary not just quantitatively, but also qualitatively (cf. Zlatev & Blomberg, 2015), proposing different mechanisms and ways in which language might influence thought. A recent example of categorization of "structural" relativities based on the type of effects of language on thought is the one proposed by Wolff and Holmes (2011), who outline a "family tree" of approaches in which they present classes and subclasses of hypothesis on how language might affect thought.[17] Within

[17] Zlatev and Blomberg (2015) propose another classification of linguistic relativity approaches that is not based on the assumption that all paradigms can be somehow located on a weak-to-strong continuum. However, at least for our purposes, the classification they offer is too broad and would not allow us to highlight the many nuances that can be individuated in current researches in linguistic relativity. One other interesting proposal in classifying studies in

the branch of theories that start from the assumption that thought is separate from language, and that language and thought differ structurally (hence avoiding a strictly deterministic view), approaches grouped in the "thinking *with* language" and "thinking *after* language" categories are the most interesting to us.

In "thinking *after* language", the long-term use of a language may direct habitual attention to specific properties of the world, even in non-linguistic context. Language can act as a spotlight, making certain aspects of the world (especially those encoded in the meanings of specific words or constructions) more salient than others. For instance, the *language-as-an-inducer* perspective suggests that a possible mechanism by which cross-linguistic differences in categorical color perception can be produced is gradual perceptual warping caused by learning. On this account, long-term experience categorizing the color spectrum using language gradually warps the perceptual representations of color resulting in more similar representations of colors grouped in the same category (i.e., those labeled by a common term) and/or less similar representations of colors grouped into distinct categories (i.e., those labeled by distinct terms). To test this it has been shown, for example, that training individuals on a new color boundary can induce categorical perception (Özgen & Davies, 2002). Many of the classical studies in linguistic relativity seem to fall under this category: Boroditsky and colleagues' (2003) study on grammatical gender, Levinson and colleagues' (e.g. Levinson, 2003) studies of spatial frames of reference (which seem to promote a particular framework for conceptualizing space but that does not imply that people are unable to use frames of reference that are not regularly encoded in their language), studies on local spatial relations (Bowerman & Choi, 2001), and on object and substances (Lucy, 1996a) (in which differences in language might bias people to attend to particular aspects of experience, such as the distinction between objects and substances). Furthermore, language may act as an inducer by priming a particular mode of processing that continues to be engaged even after language is no longer in use (Holmes & Wolff, 2010).

language and cognition is proposed by Gentner and Goldin-Meadow (2003b) based on the questions and methods that drive the various paradigms. Language can be seen as *lens*: the question posed is whether the language we acquire influences the way we see (perceive) the world. This view is close to the neo-Whorfian hypothesis that the structure of a language shapes its speakers' perception of the world. Language can be seen as a *tool kit*: the question posed is whether the language we acquire augments our capacities for representation and reasoning. Language can be seen as a *category maker*: the question posed is whether the language we acquire influences where we make our category distinction (i.e. how we conceptualize the world). This view seems closer to the Whorfian starting point of linguistic relativity.

People think *with* language when processes associated with language are activated along with nonlinguistic processes. Within this category, language can be seen as a *meddler* or as an *augmenter*. In the first case, the effects of language occur from the spontaneous recruitment of linguistic codes in tandem with nonlinguistic codes. According to the authors, this would be the case for colors, in which language can meddle with cognition via the interaction of perceptual and linguistic codes.[18] In the case of *language as meddler*, a decision can be made on the basis of either linguistic or nonlinguistic representations. In certain cases, however, linguistic representations may combine with nonlinguistic representations to enable people to perform tasks that could not be completed with either type of representation alone. Language can be seen as an *augmenter*. In this perspective, language offers a new conceptual tool, giving rise to new ways of representing and solving specific tasks. For instance, language may facilitate category learning. This type of influence was demonstrated in a set of studies by Lupyan et al. (2007), in which participants learned to distinguish between approachable and non-approachable alien creatures. These categories could be learned on the basis of visual information alone, so learning labels for the categories was not necessary for completing the task. Nevertheless, category learning was much faster when it was accompanied by auditory or written labels than not. In another work, Lupyan (2006) has shown that categories learned with labels are more resistant to interference from novel stimuli and more.

The *language-as-an-augmenter* perspective offers us a much broader framework for the study of both linguistic relativity and cognition in general. In some of his earlier works, Lupyan (2012b) argues that language, and especially the practice of labelling, can exert extremely rapid and pervasive effects on putatively non-verbal processes such as categorization, visual discrimination, and even simply detecting the presence of a stimulus. However, one puzzling issue is that effects of language seem to be both deep, in the sense that they affect even basic visual processes, and yet vulnerable, in the sense that verbal interference (disrupting access to language) can sometimes nullify those effects. For example, Winawer et al. (2007) found that when Russian-speaking subjects were placed under verbal interference, within-category comparisons no longer

[18] In addition to the studies already mentioned, relevant discoveries about how we process colors and the possible involvement and recruitment of the linguistic code in a variety of perceptual tasks are proposed by Regier and colleagues, who showed that, although there are clear universal tendencies in color naming and color perception, linguistic convention may nonetheless play a role in determining category extensions and in shaping perception that happens in the right visual field (A. L. Gilbert, Regier, Kay, & Ivry, 2006; Regier & Kay, 2009; Regier, Kay, Gilbert, & Ivry, 2010).

took longer than between-category comparisons. The transient nature of linguistic effects have been taken to downplay the role of language in cognition; it seems to reveal that the underlying structure and content of thought and perception are unaltered by palpable and general differences in language encoding (cf. Gleitman, 2010). According to this criticism, if an influence of language on, for example, color perception can be disrupted via a verbal manipulation, does this not mean that language was affecting a verbal process all along and therefore the effect is of language on language rather than language on perception? As we will discuss in more details in the following sections, such critiques are based on the assumptions that language simply maps onto concepts and on the strict separation between verbal and non-verbal processing and representations. Lupyan's *label-feedback-hypothesis* proposes another way to look at the problem (a way that will turn out to be much closer to our own understanding of the issue): language is viewed as a part of an inherently interacting system with the capacity to augment processing in a range of non-linguistic tasks by modulating on-going perceptual processes on-line. The specific progress involved in this on-line manipulation is that names (verbal labels) selectively activate perceptual features that are diagnostic of the category being labeled in our on-line processing of stimuli. An example provided by the author is the following:

> in the case of color, this means that after learning that certain colors are called "green," the perceptual representations activated by a green-colored object become warped by top-down feedback as the verbal label "green" is co-activated. This results in a temporary warping of the perceptual space with greens pushed closer together and/or greens being dragged further from non- greens. Viewing a green object becomes a hybrid visuo-linguistic experience. Knowing that some colors are called green means that our everyday experiences of seeing become affected by the verbal term, which in turn makes the visual representation more categorical (Lupyan, 2012b, p. 4).

What is key in Lupyan's proposal is that linguistic representations do not need to be understood as a separate verbal modality or channel in order for linguistic relativity effects to take place. Language influences on thought occur on-line, i.e. labels get co-activated with perceptual representations and provides top-down influences (on, for example, categorization and object perception)[19]. This perspective foreshadows some of the central argument we will be discussing

[19] Within this framework, as outlined by Zlatev and Blomberg (2015), linguistic relativity effects seem to depend on the kind of process and task being considered and, hence, it may be more productive for relativistic researches to measure the degree to which performance on specific tasks is being modulated by language, modulated differently by different languages, or is truly independent of any experimental manipulations that can be termed linguistic. However, Lupyan's perspective (from the *label-feedback-hypothesis* to *language-augmented-thought*) has much broader and relevant implications for the study of the relation between thought and language.

throughout our work, from connectionist approaches (the *feedback* works as a simple recurrent network works) to specific perspectives on the nature of linguistic meaning (i.e., words do not map onto the external world, but rather can be viewed as *cuing* meaning). Within this framework, language is taken to *interact* with other cognitive processes, not to *interfere* with them. There is a subtle but substantial difference between the two, which is largely dependent on how we define language and thoughts (see par. 1.2).

The broader framework of the *label-feedback-hypothesis* is the theory of *language-augmented thought* (Lupyan, 2012c), which is based on a precise assumption on the nature of language (and especially of words): words are not pointers to non-linguistic concepts, but are rather best described as operators on conceptual (and via continuous feedbacks, perceptual) representations. Our mental representations are to varying degrees under continuous influence of language and performance on nonverbal tasks such as categorization, visual memory, object recognition, and even simply detecting the presence of a visual stimulus is to varying degrees augmented by language. The logic to test this is rather simple: insofar as normal performance on these tasks is affected by aspects of language, manipulating linguistic variables should manipulate aspects of performance on the task. We report here the predictions proposed by the author starting from these assumptions:

> (1) Insofar as verbal labels change "nonlinguistic" representations, associating a label with a concept should affect the acquisition of the concept. Namely, labeled categories should be easier to learn than unlabeled categories. (2) Given the bidirectional information flow between the different representational layers, the effects of labels should penetrate even perceptual processes. That is, language use can actually affect what we see. (3) Named concepts should be activated differently under the on-line influence of the label than when the labels are prevented from affecting the concept (Lupyan, 2012c, pp. 261–262).

Some interesting questions emerge: what is it in labels that "augments" our representations? In what way are representations "augmented"? What cognitive mechanisms may support the interaction of language and other cognitive processes?

Learning and using labels appears to "augment" perceptual representation to make them more categorical, both when we learn a new category (Lupyan & Casasanto, 2015) and when we deploy labels to previously learned very familiar items (Lupyan & Thompson-Schill, 2012) with consequences ranging from basic perception to reasoning (Lupyan, 2012c; Lupyan & Bergen, 2016; Perry & Lupyan, 2014). As for the mechanism, the forward-feedback, bottom-up and top-

down, process envisaged by Lupyan in this work is in line with claims on the cognitive penetrability of perception (Churchland, Ramachandran, & Sejnowski, 1994; for a definition of cognitive penetrability cf. Pylyshyn, 1999) and, more specifically, with the cognitive architecture put forward with the predictive processing model of cognition (for a recent extensive outline cf. A. Clark, 2013). The question of whether language affects perception can gain clarity by considering it through the lens of predictive processing, a framework in which mental representations, from the perceptual to the cognitive, reflect an interplay between downward-flowing predictions and upward-flowing sensory signals. This framework provides a parsimonious account of how (and when) what we know ought to change what we see and helps us understand how a putatively high-level trait such as language can impact putatively low-level processes such as perception. Within this framework, language begins to take on a surprisingly central role in cognition by providing a uniquely focused and flexible means of constructing predictions against which sensory signals can be evaluated:

> [w]ords (and larger verbal constructions) become not simply ways to communicate our preexisting thoughts but highly flexible (and metabolically cheap) sources of priors throughout the neural hierarchy. This is accomplished both through flexible modification of what top-down information is brought to bear and by selectively influencing the precision weighting of prediction error, thereby influencing how much top-down information influences specific lower-level level processes (Lupyan & Clark, 2015, p. 283).

Predictive processing thus provides a plausible mechanism for many reported effects of language on perception and thought, ranging from those attested in the way language makes our perceptions more abstract and categorical to the effects of language in shaping our perception of the supposedly "same" reality.

Allow us one last remark before moving to the next topic: the language-augmented thought perspective is primarily concerned with outlining the mechanisms that underlie the influence of language on thought and with characterizing what kind of influence we can observe. In this sense, within this perspective we are momentarily setting aside the primary relativistic concern: namely, language differences. Lupyan's perspective can be viewed, at least partly, as an intra-linguistic endeavor in characterizing the nature of the influence of language on thought.[20] However, it seems

[20] If we limit our discussion to what we have reviewed so far, Lupyan's perspective is an inherently intra-linguistic approach. Recent studies, however, show a profound interest of the author for the cross-linguistic perspective and its centrality in the study of cognition (Lupyan & Lewis, 2017; B. Thompson, Roberts, Roberts, & Lupyan, 2018).

rather obvious that Lupyan's perspective is open, in line with, and supported by cross-linguistic evidence: the finding that language augments (on-line) the representation of the labeled item has clear implications for thinking about how using different languages can augment ongoing cognitive processing. Languages whose lexicons include words that refer to certain characteristics thus enable speakers (for better or worse) to selectively highlight those aspects (cf. Lupyan, 2012c).[21]

In this brief review of the current efforts in linguistic relativity research, we have tried to give a taste and sample of the variety and richness of approaches, topics, and foci that constitute the main interests of contemporary efforts. One general observation is indeed in place: the claim that the language we speak influences the way we think and perceive reality has multiple possible and equally legitimate interpretations. From anthropology, to linguistics, to psychology, it receives various kinds of explanations and it is adapted within a specific set of hypotheses about how culture, society, language, and cognition work. Hence, rather than thinking of linguistic relativity as a specific claim, we should think of linguistic relativity as a family of approaches endorsing various different interpretations of the very general underlying conviction that language is more than just for communicating thought: language plays *some* cognitive function. What the specific function of language in cognition may actually be depends on the specific interpretation we give of this underlying belief and what path any specific paradigm of research undertakes within the linguistic relativity framework. If taken together, the findings we reviewed in this and the previous section indeed point towards a positive answer to the question of the influence of language on thought, may it be in the Whorfian sense of language as the way in which we organize the "flux" of experience, or in the sense that language augments our categorical thought, or in the sense that it provides us with habitual ways of attending to reality.

[21] If you strip language diversities from the study of language and thought, would that still count as linguistic relativity? Indeed, we would not be able to look for so-called Whorfian effects, i.e. how different languages differently shape the way we think or perceive. Nonetheless, proving that language (any language) is implicated in the way we think or in the way in which we construct the world seems to be the (cognitive) basis for arguing for any kind of Whorfian effect in the first place. Linguistic relativity would then follow from the fact that languages are different. As it has been variously pointed out in the review literature of linguistic relativity (e.g. Gentner & Goldin-Meadow, 2003b; Gumperz & Levinson, 1996a), the relativity hypothesis is grounded onto two main claims: 1) languages are different; 2) the language one's speaks influences the manner in which one perceives and understands the world. Hence, different languages lead to different organizations and perceptions. Consequently, perspectives such as Lupyan's *language-augmented-cognition* can be understood as targeting the second of the two claims that ground linguistic relativity. As for the first claim, we will discuss some main issues in paragraph 1.1.3. and offer an extensive theoretical evaluation of in what way languages differ in chapter 2.

Of course, being able to properly understand the reach of the effects that language may have on thought is a central focus of current approaches in the cognitive sciences, whether they describe themselves as relativistic in nature or not. Hence, linguistic relativity (in its various forms) can be viewed as one of the possible answers to the more general question about the cognitive function of language (cf. Carruthers, 2002). If understood within this broader framework, linguistic relativity elicits a different (definitely less provocative) vibe and a much broader quest, constantly reshaped and integrated with newly acquired knowledge about how our brain and our cognition might work more broadly. In the next chapters, we will see how the study of language in cognition has found and still finds a multitude of approaches interested in many different aspects of this relation. The question of language and thought is not one question, but several; "[w]hether language has an impact on thought depends, of course, on how we define language and how we define thought. But it also depends on what we take to be the criterion for 'having an impact on' " (Gentner & Goldin-Meadow, 2003b, p. 12). Before turning to a discussion of the assumptions that underlie linguistic relativity approaches, let us briefly review some of the main reasons why the debate around linguistic relativity is far from being settled.

1.1.3. A debate far from being settled

As we already had the chance to sketch in the previous sections, linguistic relativity was not always welcome by researches in the cognitive sciences. There are a few main reasons why linguistic relativity failed to gain popularity and traction, some of which are intimately connected with the rise of the Generative paradigm. In many cases, the relevant studies produced were susceptible to valid methodological criticisms; in some other cases, the opposition was grounded on a more theoretical basis. Some of the classical criticisms to linguistic relativity have definitely been overcome as many resulted from misperceptions of some relativistic claims. Other criticisms are indeed still valid and need to be specifically addressed. Hence, the debate on whether and to what extent linguistic relativity is a compelling claim is far from being resolved and it can be viewed as comprising a set of major issues[22]: in the first place, researches have long debated on

[22] In very general terms, the broader framework that gave rise to criticisms to the linguistic relativity hypothesis is the debate between universalism and relativism, which comprises many different issues on many different levels. In the first place, as we will see, there is the question whether language is universal or if different languages actually show considerable variability. In the second place, the universalism versus relativism debate runs parallel to the nature

the very issue of language diversities, i.e. if languages are actually so diverse as the linguistic relativity hypothesis claims; in the second place, linguistic relativity was often charged with language determinism, which is indeed conceptually unsound; in the third place, in order to show relativistic effects, relativistic approaches need to prove that language can be separated from both thought and culture. We will look at these issues in turn.

The first premise of the linguistic relativity hypothesis is that languages are different. Our intuition tells us that this may indeed be a valid claim to begin with; a very coarse example of this diversity is that when we have to learn a new language, especially during our adult life, we are faced with the fact that the sounds may be different, the grammar is most likely to show substantial diversities, as well as word order in the way the sentence is constructed, up to the whole set of new words we have to learn in order to be able to say or understand anything at all. Is this enough to say that languages are different? Not really. In order for linguistic relativity to have a solid ground, languages should exhibit a variation that is not merely superficial, but instead structural. The words we learn when studying a new language may indeed just be different labels for the same conceptual representation or for describing the same facts in reality. Moreover, even if languages differ in the way they build a sentence, it may well be the case that the underlying grammar and structure of languages at a deeper level show instead substantial universality.

Universalist claims about language constitute a strong argument against the very possibility of linguistic relativity: a widespread assumption among cognitive scientists, growing out of the generative tradition in linguistics, is that all languages are English-like, but with different sound systems and vocabularies. Differences between languages are merely superficial, and they can be resolved by postulating a more abstract formal level at which individual language differences disappear. This universalist claim about language became so popular that "it now pervades a great deal of work done in psycholinguistics, in theories of language evolution, language acquisition, neurocognition, parsing and speech recognition, and just about every branch of the cognitive sciences" (N. Evans & Levinson, 2009, p. 429). If this claim was to be true, then linguistic relativity would have no basis whatsoever.

However, decades of research in linguistics, anthropology, sociology, and so on, have

versus nurture debate, i.e. the ever-lasting question of whether we owe what we know and who we are mostly to innate processes and contents or to experientially, socially, culturally, linguistically learned contents. We decided not to frame our debate on linguistic relativity in these terms given their overarching reach, which would have taken us far from the aim of our work. Instead, we will try to address more specific issues.

showed that the myth of language universals is contradicted by a large amount of evidence (for a review of the study and a useful discussion cf. V. Evans, 2014). Languages exhibit profound differences at all levels of organization, from the number and type of sounds, to the kind of morphology, to the way in which they carve nature to its joints. In Evans and Levinson's words: "[t]he fact is that it is a jungle out there: languages differ in fundamental ways – in their sound systems (even whether they have one), in their grammar, and in their semantics" (N. Evans & Levinson, 2009, p. 438).

Drawing from the review of studies presented above, the most relevant aspect of this diversity, in order to establish the place of linguistic relativity in the study of language and cognition, is that languages are not entirely neutral with respect to the information they carry. A careful study of cross-linguistic differences reveals that every language reflects a certain perspective on the world through its inventory of words and encoding strategies: word meanings sometimes work against the correlational structure of the world and diversity in word meanings across languages is pervasive (Wolff & Malt, 2010). These claims, which are often taken to be a novelty from the point of view of the cognitive sciences, were exactly the observational starting point of Whorf (and of many others before him): languages differ in their semantic partitioning of the world. For example, words tend to vary in how many distinctions within a domain are encoded in language, i.e. languages differ markedly in how they partition by name many domains including, as we have shown, color, space, body parts, but also motion, emotion, mental states, ordinary household containers, and so on. Languages show profound variations in the way they partition both nature, carving it up at its joints as much as reflecting them, as well as our conceptual space. Words cannot be taken as mirroring neither the organization of nature nor the organization of our conceptual structure. This will be the main focus of our discussion in Chapter 2. As for now, given the extent of documented diversities, it seems safe to state that languages differ not just at a superficial level but indeed in their very structure and in the kind of information they convey.

This brings us to the second point of our discussion: linguistic relativity is not, and cannot be, linguistic determinism. We are here thrown at the heart of the language and cognition debate. Is it the same to argue that language influences thought, or that language determines thought, or that language is the same as thought? Obviously not, and linguistic relativity (in its various forms) falls indeed within the first claim, both in its theoretical arguments and in its methodological formulations. In the first place, the idea that conceptual structure and semantic structure are closely

coupled has been the dominant position in traditional cognitive sciences, more than in the relativist arena (Gentner & Goldin-Meadow, 2003b). Within cognitive psychology, there was a strong sense that concepts come first, and that language merely names them; human conceptual structure is relatively constant in its core features across cultures, and conceptual structure in merely mirrored by semantic structure. The view from linguistic relativity is precisely the opposite: language and non-linguistic cognition are not the same thing, words are not the units of thought, and thought and language exhibit different structures, so much so that language is taken to interfere with our non-linguistic knowledge of the world. Hence, Pinker's (1994) "obituary" for linguistic relativity, in which he argues that "the idea that language shapes thinking seemed plausible when scientists were in the dark about how thinking works or even how to study it" (1994, p. 58) is due to a misunderstanding on the actual point made by linguistic relativity: "[t]he idea that thought is the same thing as language" (1994, p. 51). In other words, Pinker equates the latter Orwellian claim that "we think in language" (see discussion in Casasanto, 2008) with the relativistic position that "language shapes thought". This caricaturization allows for a facile assault of the relativistic position. It is clear, after all, that thought is not isomorphic with language, which is a position at the very heart of the relativistic claim. As we have seen in the previous section, linguistic relativity aims to uncover the mechanisms of how language influences thought mostly starting from the assumption that linguistic influences can be observed in that way they shape non-linguistic cognition. As we will better discuss in the next section, the fundamental idea of linguistic relativity is that the two, language and thought, have to be kept separate and can be separately investigated.

Hence, given that the equation of language and thought is definitely far from being a relativistic assumption, the next objection has to do with the strength of the influence: linguistic determinism, outside of the language and thought equation, is taken to imply that the shaping role of language is so strong that it can even overwrite pre-existing perceptual and conceptual capabilities (Gentner & Boroditsky, 2001; Lucy & Gaskin, 2001; cf.: Bowerman & Levinson, 2001; Wolff & Holmes, 2011).[23] If it might as well be true, in some specific instances of some domains, that language determines thought so deeply that, for example, infants lose their ability to

[23] Not even Whorf argued for such a strong version of linguistic relativity: "[t]he tremendous importance of language cannot, in my opinion, be taken to mean necessarily that nothing is back of it of the nature of what has traditionally been called 'mind'. My own studies suggest, to me, that language, for all its kingly role, is in some sense a superficial embroidery upon deeper processes of consciousness, which are necessary before any communication, signaling, or symbolism whatsoever can occur" (Whorf, 1956, p. 239).

notice phonetic distinctions absent in their native language (e.g. Werker & Tees, 1984), in many other cases a looser connection between language and thought is revealed from cross-linguistic studies. Similarly, as we have seen in the previous section, linguistic biases can be disrupted when we interfere with language. Even if language is not shown to strictly determine thought, this does not imply that linguistic relativity is at loss. As Wolff and Holmes argue: "[t]he lack of alignment between language and thought raises severe problems for linguistic determinism, but these problems do not imply that language cannot have an effect on thought. Indeed, it is because language and the conceptual system differ that we might expect a tension between them, driving each system to exert an influence on the other" (2011, p. 255). To put it differently: accepting the existence of significant linguistic diversities, along with the associated cognitive effects of such diversities, does not imply that universal aspects of human cognition need to be completely ruled out. Most contemporary researchers (but we have seen that a similar perspective was already present in Boas, Sapir, and Whorf's works) presume a fair amount of universalism in human cognition. However, the extent of that universalism (as well as the extent of variation) is an empirical matter that requires exploration. Proponents of linguistic relativity believe that this exploration will allow us to better establish which facets of human cognition are in fact susceptible to variance, perhaps due to linguistic influence.

The last criticism that we aim to review here is concerned with the very possibility of isolating language from culture more broadly or from the rest of our cognition. One of the issues that research on relativity has to face is the challenge of extricating linguistic influences from nonlinguistic cultural influences on cognition. In conducting such work, it may be misleading to demonstrate that cognition varies across two tested populations, and then offer a strictly linguistic interpretation of such variation. After all, other nonlinguistic cultural variables may be at play in such variation. In short, the anti-relativists often accurately claim, sometimes implicitly, that correlation does not imply linguistic causation. In *The Stuff of Thought* (2007), Pinker argues that just because languages express ideas in different ways (or even express ideas that other languages have not thought of expressing) it does not follow that language influences, and moreover fundamentally affects, thought and perception. His point is that the mere fact that different groups of speakers behave in different ways from others, and also, coincidentally, speak a language that is organized in a different way, does not mean that language has caused the behavior. Within the framework of this objection, the majority of the criticisms stem from arguing against particular

instances of the relativistic methodology, which are taken not to be sophisticated enough to isolate linguistic aspects from broader cultural aspects (or, for that matter, from thought itself). The problem of causation highlights two major possible pitfalls of the relativistic enterprise: that of separating language from culture, and the problems faced with disentangling language from thought. In general, behavioral tasks (on which the relativistic enterprise in mostly based) merely allow us to infer an underlying cause and, hence, cannot be taken as demonstrating causation properly; do to so, they have to be supplemented with other methods.

On the side of language and culture (which we will not cover extensively here since we are mostly focused on the language and thought relationship), if it might be fairly easy to demonstrate that there are variations in behaviors across cultures, less straightforward is to tie these variations specifically to linguistic differences between the groups. This problem has been variously addressed in the relativistic literature. We will here report some of the main attempts that have been made to clarify the relativistic procedures, since they shed light to some of the main diagnostic procedures that underlie the study of linguistic relativity.[24]

Lucy discusses this issue directly: the demonstration of the decisive role of language must come from "controlled evidence capable of distinguishing among competing accounts of a given pattern of cultural thought and behavior" (1992, p. 270). To this regard, experimental psychology seems to provide a possible line of solution: experimental controls are introduced precisely to distinguish among various possible influences on behavior and thereby clarify functional interrelations. For example, task behaviors designed to predicted effects of language patterns can be made very specific and highly characteristic. This reduces the likelihood that any simple alternative explanation may be suitable to predict the same highly characteristic pattern of results. Moreover, one may test speakers from populations that are very similar culturally, and even linguistically, apart from one particularly relevant variable (e.g. Pederson, 1995). As we have mentioned before, the logic underpinning Lupyan's experimental work does not diverge much: insofar as normal performance on some task is affected by specific patterns of language, manipulating linguistic variables should manipulate predicable aspects of performance on the task

[24] Establishing the place of language within a broader cultural framework is definitely beyond the scope of this introduction and beyond the scope of our work in general. However, even if we are not primarily concerned with how language can be disentangled from culture and with what this might mean for the study of linguistic relativity, it seems important to at least sketch the arguments that have been proposed to this regard, even if we are not offering a critic discussion of them.

(Lupyan, 2012c). Furthermore, bearing in mind that correlation is not causation, Everett points out that "it is worth noting in this discussion of issues with some anti-relativist positions (more specifically, those that are recalcitrant vis-à-vis the existence of any linguistic influences on thought), that extensive correlation of the sort in question also cannot be ignored. We cannot discard inductive reasoning in our attempts to better understand these issues, after all. If numerous correlations are established between particular linguistic features and particular behaviors, for example directional term types and spatial orientation strategies, the consistency of such correlations may crucially inform hypotheses on the subject. Correlation may not establish causation definitively, but it often points to the most reasonable explanation of a series of observations" (2013, p. 31). Hence, it may be the case that repeated and controlled experimental procedures allow us to actually point to language specifically as the source of the deviation observed in behavioral tasks.

If separating language effects from more broadly cultural influences is (at least in principle) methodologically possible, what about disentangling language from thought in the first place? As we have pointed out multiple times throughout our discussion so far, the neo-Whorfian perspective aims at addressing the influence of linguistic patterns and categories on non-linguistic aspects of cognition; hence, the separability of the two is fundamental in order to study the influence of language on cognition. Defining language and thought in a way that allows them to be both distinguished and co-related is a first prerequisite for further investigations into their relationship from a relativistic perspective. In the next sections we will discuss this issue as well as some of the main problems in which we might fall into when trying to separate language and thought.

1.2. *Rethinking the assumptions: language and thought*

1.2.1. Language, thought, and reality

In his 1992 masterpiece, Lucy points out some fundamental methodological issues in the study of linguistic relativity. Using Whorf's formulation as a starting point and talking into account the general course of previous research, he extracts three key dimensions of the relativity question: the first two are, of course, language and thought, the point being the study of the influence on the first on the latter. The presence of these two dimensions gives rise to a first methodological issue:

> it is important that there be a clear analytic distinction between linguistic categories and cognitive categories so that the influence of the former on the latter (if any) can be detected and identified. If linguistic and cognitive categories are conflated in the research formulation, then the findings will be ambiguous; if they are directly equated with one another, then what is to be proven will, in effect, have been presupposed (Lucy, 1992, p. 264).

Hence, the distinction between linguistic and cognitive categories is the prerequisite of any relativistic inquiry. How can we distinguish between the two, both in terms of theory and in terms of methods? Especially in the psycholinguistic literature on relativity, the answer to this question is: by means of taking reality as the anchor point. Reality, the third key dimension of the linguistic relativity hypothesis, is taken to be at least analytically independent of both language and thought, and it is invoked to anchor the comparison of diverse language categories and to bridge the analyses of language and thought. In Lucy's words: "in this American tradition, the research focus has been on the way diverse language categories implicitly classify reality and then on how those implicit classifications, in turn, influence thought about or the view of reality of those who speak them. And a crucial theoretical-methodological problem (whether recognized or not) has been to characterize reality without simply invoking the conception of reality suggested by the investigator's own language" (Lucy, 1992, p. 264).

It seems clear then that the relativity question sets itself precisely at the intersection of these three dimensions: language, thought, and reality. The language system, the conceptual system, and the world are the three poles around which the debate has been oriented. As we have seen, reality is taken to be independent from any of the other two terms and can be used as a means of

comparison. The language system and the conceptual systems are different. The way to establish the differences between them, and hence to set up any instance of the relativistic inquiry, is to compare the language system and the conceptual system against a common portion of reality. This essentially means, as we will see in more details in the next section, that we can select a certain domain of reality, evaluate separately how language and thought encode it, and then manipulate the linguistic input in the empirical studies to estimate the degree to which language is involved in a given cognitive process.

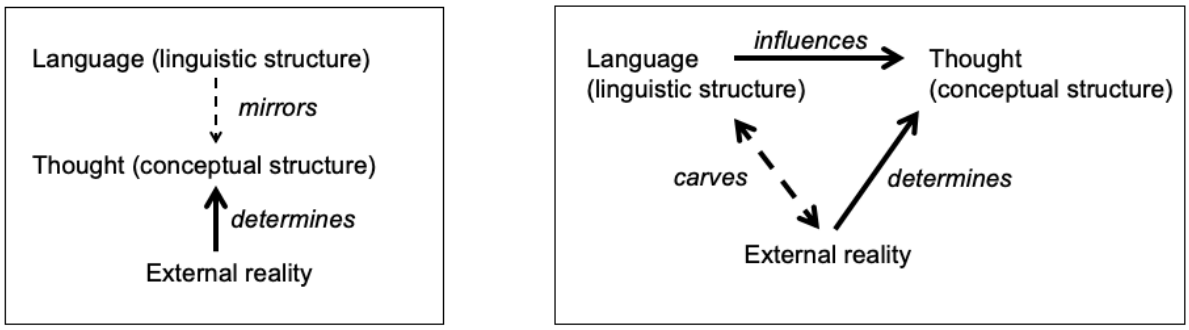
Although this may seem in principle a very straightforward method, it hides multiple theoretical as well as methodological obstacles. In the first place, as Lucy points out, describing reality in a neutral, unbiased way is far from being simple and obvious. In fact, even in those domains in which a description based on physical and biological investigations may give us an objective lead in partitioning reality, we observe that often the conception of reality of the investigator's own language plays a non-neglectable role in comparing different languages. Berlin and Kay's procedure in discovering basic color terms seems to be a case in point. For example, Evans (2014) points out that, in respect to the supposed priority and evolutionary sequence of basic color terms, the neo-Whorfians (e.g. Lucy, 1997; Saunders, 2000) have made four main criticisms of this research tradition as it relates to linguistic relativity; the first criticisms is that theoretical construct of the 'basic color terms' is based on English, and hence that it does not provide a completely language-proof access to reality: "[i]t is then assumed that basic colour terms - based on English - correspond to an innate biological specification. But the assumption that basic colour terms - based on English - correspond to universal semantic constraints, due to our common biology, biases the findings in advance. The 'finding' that other languages also have basic colour terms is a consequence of a self-fulfilling prophecy: as English has been 'found' to exhibit basic colour terms, all other languages will too" (V. Evans, 2014, p. 204). The point here is not as much about whether color categories are universal or not, but rather that the very method that should allow us (and has been long taken) to establish differences between languages in the first place, as well as our possibly innate color sense and its effect on language, is often biased by the language spoken by the investigators (which is generally English). On the contrary, a neutral description of reality for the purposes of comparisons should not privilege the categories of any one language or

culture at the outset and should not be reduced to the categories of a particular language or system of thought. [25]

In more theoretical terms, it seems that, in assuming a relativistic perspective, we are also already assuming a set of assumptions about the nature of meaning in language and in cognition, and especially about the kinds of relationships between the three dimensions. Hence, if we take linguistic relativity to be one possible answer to the much broader question of the relationship between language and thought, then it seems that research in linguistic relativity assumes what instead it is supposed to investigate: if one necessary assumption is the distinction between language and thought, then linguistic relativity, by setting off assuming the separability of language and thought, is taking a preliminary position on the relationships between the two. In other words, if one of the outcomes of the linguistic relativity enterprise is to establish which relationship is entailed between language and thought (language influences thought), but in order to study it we already have to assume that the two are separable in an identifiable way (the two are structurally different), then the assumption itself is already part of the answer. In order to avoid the ghost of circularity embedded in such claims, reality is presented as the third (independent) element that, as we have seen, should allow us to establish the differences between the two terms without needing to investigate the relationship between them at the outset.

Thus, it seems that the first step into grounding the study of linguistic relativity would be to determine what kind of connection there is between language and reality and between thought and reality. Researches from the cognitive sciences, as we will see in more details in chapter 2, suggest that the connection between thought and the world is tighter than the connection between thought and language. Hence, thought is determined by the organization of the world in relation to the nature of our cognitive abilities, and language simply maps onto already established conceptual distinctions (this view is generally addressed as holding a priority of the cognitive over the

[25] Lucy proposed extensive cross-linguistic comparison as a method to provide a neutral description of reality, by achieving a descriptive *metalanguage*. "One essential step in building such a descriptive language is to develop it from a *broad comparative sample* of languages and cultures so that it can reflect a maximum range of diversity at the outset and place each specific system relative to the array of possibilities. [...] the most promising approach at present is to turn to contrastive linguistics for help in developing a provisional, neutral description of reality. In other words, *reality can be described as it appears through the window of language*. The description will be neutral to the extent that it succeeds in not favoring any one language, but it will remain a decidedly linguistic vision of the world and, in this respect, a decidedly partial construal of reality" (Lucy, 1992, pp. 274–275). This method finds its empirical application in *Grammatical Categories and Cognition* (1996a) in the evaluation of how languages encode numerical attributes. Apart from Lucy's work, this method did not find substantial applications in the psycholinguistic literature that mostly remained tied to the approach described above.



A. The standard view from the cognitive sciences

B. The view from linguistic relativity

Figure 1. Schematic representation of two possible views on the relationships between language, thought, and reality.

linguistic). Since the level of the linguistic is essentially thought of as mapping onto our conceptual structure, little or no effort has been devoted within the cognitive sciences to evaluate the relationship between language and the world (cf. Wolff & Holmes, 2011). Language is generally taken as a window onto our conceptual structure, i.e. it simply reveals how our concepts are organized, and the nature of this organization is taken to be dependent on our relationship with the external world. In contrast, linguistic relativity holds that there may be a connection between language and thought in which language influences thought while still maintaining a formative connection between thought and the world (see figure 1 for a schematic representation). Within this view, language has a direct connection to reality and also to the realm of thought, which, as we have seen before, depends both on how we interact with the external world and on what language we speak. Supported by evidence from cross-linguistic differences in the way languages encode the external reality, linguistic relativity researchers view language as both *being carved by the joints of nature* as well as *carving nature at its joints*. This means that language does not simply reflect the organization of the external world but, as we have seen before, it rather selects what aspects of the world are being represented and where to set the boundaries between linguistic categories.

Let us now take a closer look at how this general framework unfolds empirically and methodologically in the study of linguistic relativity. As it will become clear in the next sections, our conclusions about the possibility of separating language and thought by means of comparing them on how they encode reality are not particularly encouraging. This is due to what we will argue are some main limitations in how language and linguistic meaning are theorized within the

relativistic framework: on the one side, the lesson from cognitive linguistics shows us that language and thought cannot be taken as completely independent one from the other. Language and the mind cannot be understood without and outside one another; treating language and thought as fundamentally different "modules" or "representations" is mistaken (Lupyan, 2012c; for a discussion cf. Zlatev & Blomberg, 2015). On the other side, linguistic relativity has mostly focused on the referential function of language leaving many other aspects of linguistic meaning almost completely unexplored.

1.2.2. Language and reality

Let us now take a slightly different perspective on the issue and briefly consider the domains that have been the focus of the relativistic enterprise. It may have become clear at this point that the majority of the studies and approaches reviewed so far are interested in how language influences our perception or categorization of concrete entities that can be individuated in the external world, both in inter-linguistic and intra-linguistic studies, for which we can (more or less) easily individuate non-linguistic referents. The theoretical reason that underlies this selection of domains is, as we have seen, the need to establish differences between language and thought based on the comparison with the external world. Hence, one practical consequence is to pick out domains that can be described both in linguistic and non-linguistic terms (the more objective, the better).

The domain of color is the epitome of such an approach. In fact, having handy the physically-based organization of the color spectrum (that, in theory, provides us with the scientific evidence of what colors actually are and where the boundaries of a given color should be placed in terms of wavelengths, frequency, or hue), we can compare different languages in where they place the relevant boundaries between categories. For example, for colors of wavelengths between 492nm and 455nm, some languages may distinguish between *blu* and *azzurro*, whereas some others may not. We can then use specific instances of these colors (e.g. patches of color) as inputs to our studies, keeping in mind how they are categorized in language, to evaluate whether the categorization operated by language has an impact on what we see or on what we think. Similarly, in Majid's (2010) study on how different language partition body parts, one of the tasks was to color in a selection of body part terms from participants own language (cf. van Staden & Majid, 2006). Similarly, in studying the effect of language on categorization, Lupyan draws on the

fundamental property of language of referring to objects with labels (e.g. using to word *chair* to refer to a chair), and suggests that verbal labels do not simply point to nonlinguistic concepts (e.g. the word *chair* does not simply map onto the concept of "chair"), but rather actively modulate object representations that are brought on-line during (supposedly) non-verbal tasks (Lupyan, 2012c). Besides the theoretical aspects, dealing with concrete domains with easily identifiable referents is much more practical in terms of experimental setup. Very intuitively, testing the role of words such as *democracy* or *freedom* on cognitive representations is not as straightforward as testing the role of *blue* or *goluboy* in how we perceive the color that those words label.

Of the various components of language, the referential and denotational value of words has generally been the main target of research interests. The presuppositional relation of words to immediate external and contextual features such as objects and events provides a crucial anchor (or common reference point) for linguistic comparison and for non-linguistic assessment of thought. Our analysis so far has brought us close to what we think is at the heart of any inquiry into the relationship between language and thought: each and every proposal about the nature of this relationship is shaped by the assumed definition of meaning. For instance, studies in linguistic relativity tend to isolate the referential component of meaning, often without considering the meaning of single lexical units within the broader structure offered by language. Psychologists often speak about language and meaning in terms of individual words that label things in the world. Even within the relativistic framework, in which words do not simply map onto the external reality but select what aspects of reality to represent, the study of language and linguistic meaning is often limited to what referents a given word or set of words denote. Words are often conceived of as self-sufficient forms with straightforward meanings merely derivative of experience.

In contrast, as we will further argue in the next chapter, even referential word meanings depend heavily on the place that the given unit occupies in the system of language. The acknowledgment of the systematic aspects of linguistic meaning is often left out of the study of the relationship between language and thought; one outstanding consequence is that the failure to recognize the structure-dependent quality of word meanings makes it difficult to understand their dynamic nature and the possible reach of the effects that the structure of language may have on cognition. In the course of our work we will argue and (hopefully) show that a structure-oriented understanding of linguistic meaning (i.e. linguistic meaning is not just the referential/denotation aspect of words) offers important insights into the full potential for language to influence thinking.

Interestingly, the structural aspects of meaning that cut across the language system were the very starting point of Whorf's argument and the offspring of the linguistic relativity principle. As we have seen, Whorf largely focused on language as a system of *fashions of speaking* (whole configurations of meanings) both theoretically and from an observational point of view; these fashions of speaking cut through the language system and comprise overt and covert categories by means of which language organizes experience. And even when focusing on single words or patterns limited in range, Whorf was particularly careful in highlighting the need for a structural and grammatical discussion of meaning:

> The linguistic material in the above examples is limited to single words, phrases, and patterns of limited range. One cannot study the behavioral compulsiveness of such material without suspecting a much more far-reaching compulsion from large-scale patterning of grammatical categories, such as plurality, gender, and similar classifications (animate, inanimate, etc.), tenses, voices, and other verb forms, classifications of the type "parts of speech," and the matter of whether a given experience is denoted by a unit morpheme, an inflected word, or a syntactical combination (Whorf, 1956, p. 137).

Following Whorf's view, human languages as particular systems of meaning can guide speakers into one direction of thought rather than another.

> Concepts depend upon the nature of the language or languages through the use of which they have been developed. They do not depend so much upon any one system (e.g., tense, or nouns) within the grammar as upon the ways of analyzing and reporting experience which have become fixed in the language as an integrated "fashion of speaking" and which cut across the typical grammatical classifications, so that such a "fashion" may include lexical, morphological, syntactic, and otherwise systemically diverse means coordinated in a certain frame of consistency (Whorf, 1956, p. 158)

Our concepts are shaped by very large-scale structures of meaning that draw on the cross-cut multiple systems in language. Hence, subsequent work, seeking to evaluate Whorf's ideas empirically by research on individual words or small sets of words without reference to their place in the language as a whole, misses one of the central insights of his claim.

And even if we set Whorf's perspective aside in the study of the relation between language and thought, we are still confronted with a long tradition in linguistics that contributed in highlighting the relational and differential nature of linguistic meaning. The meaning of a lexeme,

even when and if it has denotational referential value, arises also from the place of the unit within a broader set of relationships internal to the system of language. In the Saussurian and structural tradition, structural meaning arises both paradigmatically, from its selection from an array of available forms with contrasting values, and syntagmatically, from its placement with other units in utterances. Researches in linguistic relativity have often overlooked this aspect of linguistic meaning, circumventing or minimizing attention to structure in several ways: for example, by focusing of words in isolations without any reference to their grammatical or other systematic properties, or by attending to grammatical structure only in a very selective way (for a review cf. Lucy, 2010).

One outstanding exception to this general oversight of the structural aspects of meaning is Lucy's work on numeral markers.[26] Lucy directly recognizes the need to focus on the structural aspects of meaning:

> [s]tructural patterns are ubiquitous, leading speakers to monitor constantly (or habitually) certain relevant features of meaning across a wide array of forms. They are often obligatory, in which case speakers are actually compelled to attend to and signal them. And they are typically systematic, such that the various components of structural meaning are interrelated, leading to cross-cutting entailments among semantic categories that reinforce and interact with one another (Lucy, 2010, pp. 272–273).

Drawing from the works of Whorf and Vygotsky, in Lucy's research the structure of language is re-established as "*the key element of the language and thought relationship*" (Lucy, 2010, p. 273). In practical terms, Lucy sets off with an in-depth analysis of number marking semantics in American English and Yucatec Maya, analyzing the grammatical differences between the two languages and how they correspond do differences in lexical semantics having to do with quantification (Lucy, 1996a). In a recent summary of his main empirical work, Lucy draws the following conclusion:

> The crucial point in the present context is that apparently similar forms in the two languages can have very different underlying lexical meanings. More specifically, denotational overlap, that is, having two forms that select the same object in the world in a given context, does not amount to semantic equivalence, that is, haying the same lexical meaning. In context, both *kib'* and *candle* can refer to candles, but *kib'* accomplishes this by reference to the material and *candle* by reference to other

[26] For an in-depth discussion of Lucy's work we refer to our previous work (De Luca, 2017).

> characteristics such as shape, composition, and use. Crucially, this meaning difference is not an isolated, idiosyncratic one, characteristic only of this translation pair, but rather forms part of a systematic pattern of lexical meaning for all nouns, which in turn forms part of an overall number-marking schema of plural and enumeration in the language (Lucy, 2010, pp. 276–277).

Hence, if we assume a Saussurian perspective, two conclusions are here possible: in the first place, although advocating for the centrality of the structural aspects of language, Lucy's work, of the various components of grammar, essentially targets referential categories with clear denotational value (cf. Lucy, 1992). Within this perspective, it is clear that languages exhibit differences in the morphosyntactic structures available for acts of referring. The linguistic relativity hypothesis then asserts that these differences in morphosyntactic structure have detectable effects on thought about reality. In the second place, reducing structural aspects of language to grammatical and morphosyntactic patterns still does not allow us to uncover the properly structural aspects of linguistic meaning. In the rest of our work, drawing from the Saussurian notion of meaning and of language as a system, we will try to outline a possible way in which research interested in the relation between language and thought might benefit from a more linguistically informed definition of meaning. We will also try to focus on the consequences that such a definition of meaning, together with recent developments in cognitive linguistics, might have for a general understanding of the role of language in cognition. Before turning to a more specific layout of our proposal, we want to reconsider in more details the two fundamental mechanisms that have been but forwards for explaining the interaction of language and thought; we will propose some general observations on the entailments, and differences in entailments, of such positions.

1.2.3. Language and thought: *interference* and *ingredient*

If we assume a relativistic perspective on the study of language and thought, words and concepts need to be kept separated from both a methodological and an ontological perspective (Zlatev & Blomberg, 2015), i.e. meaning in language and meaning in cognition cannot be straightforwardly equated. The main tenant of the relativistic enterprise is that language provides us with a set of representations that somehow interacts with our otherwise non-linguistic cognitive abilities, hence modifying our non-linguistic behavior. In the previous section, we have pointed out and argued that languages do not simply mirror the organization of the external world, but rather (at least partially) determine the structure and organization of the outside reality in a

language-dependent manner. In turn, this organization and structure is taken to be different from the conceptual one and, hence, the actual source of influence on cognitive processes.

A first question that spontaneously arises is the following: if we take the structure of language to be different from the structure of the external world and from cognition, where does language gets its structure from? The second question regards the mechanism of the interaction of language and thought: if language provides us with a content different from that of sensory experience and cognition, how do the two interact? The answers to these two questions go far beyond the claims made by relativistic approaches to meaning (and this is indeed not just a limitation of the relativistic framework) but, as we have tried to show above, it is nonetheless necessary to attempt to delineate a solid perspective that would allow us to better frame the debate on the nature of the relationship between language and thought, supported by advancements in both linguistics and in the cognitive sciences. We are of course not claiming that we have a solution or an answer to these questions: our overall goal is just to lay out a possible line of inquiry on these issues at the intersection of linguistics and the cognitive sciences. Let us begin with a quick overview of the first of the two questions outlined above.

Where does language gets its structure from? Even if we assume that language and thought are two separate things providing us with different contents, language is inherently (at least in part) a cognitive phenomenon and, hence, it falls under the domain of psychological inquiries. The attitude of psychologists towards linguistics, as well as that of linguists towards psychology, has been and still is indeed problematic, making any effort of combing perspectives from the two views puzzling to say the least. One fundamental issue that underlies any relativist claim has been that of combining a somehow independent view of language and linguistic meaning (i.e. language does not simply mirror the external reality or the structure of thought) with the undeniable fact that language stems from and is indeed part of cognition.

Even outside of the relativistic frameworks, there is a range of views on the relationship between the abstract (to some extent independent) system of language and how language is used; one common view that became popular with the first cognitive revolution, especially thanks to the Chomskian paradigm, is that the system of language has an essential and unique internal structure and what people say is a potentially imperfect reflection of this inner essence, muddied by performance factors. From a cognitive perspective, one of the main consequences was to encapsulate the structure of language is a specific and independent module in cognition. In recent

years, new approaches have been developing an opposite view: it is argued that the principles needed to explain linguistic behavior are domain-general and based on the impact that specific experiences have on the mental organization and representation of language (for a collection of works by leading proponents of the cognitive-functional approach to language cf. Tomasello, 1998b). Linguistic structure can be derived from the application of domain-general processes as they operate in multiple instances of language use, as for example categorization, chunking, rich memory storage, analogy and cross-modal association (Bybee, 2010). Hence, domain-general cognitive abilities as well as more specific contents we extract from our (perceptual and physical) interaction with the world become fundamental in shaping language itself. Furthermore, the second cognitive revolution and the subsequent studies in cognitive semantics and in cognitive approaches to grammar contributed in highlighting the encyclopedic nature of linguistic meaning: lexical concepts do not represent neatly packaged bundles of meaning. Rather, they serve as 'points of access' to vast repositories of knowledge relating to a particular concept or conceptual domain (e.g. Langacker, 1987). Accordingly, meaning is constructed at the conceptual level. Meaning construction is equated with conceptualization, a process whereby linguistic units serve as prompts for an array of conceptual operations and the recruitment of background knowledge. Meaning is a process rather than a discrete 'thing' that can be 'packaged' by language (e.g. Langacker, 2008; for a review cf. V. Evans, Bergen, & Zinken, 2007). Hence, since language arises from non-language specific cognitive abilities, meaning is inherently encyclopedic, and language allows us to communicate thought, by examining the nature of language "we gain a direct vista into the nature and mechanics of the human conceptual system, upon which language depends for its semantic prowess. Language reveals the structure of thought: it is a window on the mind" (V. Evans, 2015a, p. 18). Although favoring a view in which words can and do provide clues as to the nature of concepts, Evans explicitly recognized (and builds upon) evidences that speak for the power of language to transform minds in different ways (i.e. the linguistic relativity hypothesis): "[t]he relationship between language and the mind is a dynamic one: while language can serve as a window on the mind, it also plays a role in shaping the way we think" (V. Evans, 2015a, p. 20).

Furthermore, the traditional dichotomy between language and thought could be misleading in another way. As Boroditsky (2011) notes: [l]anguage also appears to be involved in many more aspects of our mental lives than scientists had previously supposed. People rely on language even when doing simple things like distinguishing patches of color, counting dots on a screen or

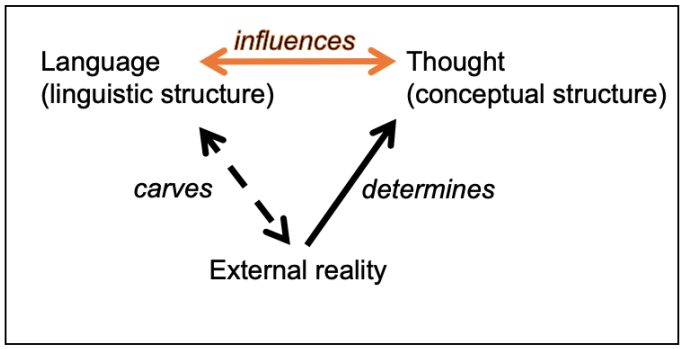


Figure 2. Rethinking the relativistic assumptions: the direction of the influence between language and thought is inherently bidirectional.

orienting in a small room [...]. What researchers have been calling "thinking" this whole time actually appears to be a collection of both linguistic and nonlinguistic processes. As a result, there may not be a lot of adult human thinking where language does not play a role" (2011, p. 65). The extent to which language is involved in cognition could go well beyond what hypothesized by relativists if, for example, it turns out that language-related brain areas are activated not only when we process language stimuli but in many other non-linguistic situations. And, more generally, it might as well be the case that no neat line of distinction can be traced between linguistic and non-linguistic cognitive processes. Rather than considering the ways in which language effects thought, we could frame an entire discussion along different lines, for example: How much of thought is linguistic? How closely intertwined are linguistic thoughts and other types of thoughts? Are linguistic facets of thought extricable from the greater web of thought? Without going deep into this discussion, which if far beyond our reach, our point is that the influence of language on thought should be understood as a bidirectional relationship. Language influences thought as much as is influenced by domain-general cognitive processes and non-linguistic cognitive content (see figure 2)[27]. Researches in linguistic relativity by no means deny it. To some researchers at least, language represents one of many types of thought, and when we explore linguistic relativity we are essentially exploring the extent to which this type of thought interacts with others.

[27] Throughout the discussion presented so far, we have been carefully avoiding considering the relationship between thought and the world. As it will become clear in the rest of our work, if we aim at establishing the differences between language and thought, we need to engage with how thoughts may come into being in a non-linguistic framework, how they are structured, where they take their structure from, and so on. However, it is clear that an in-depth discussion of the nature of the relationship between thought and the world would imply a completely different type of work. Hence, for the purposes of this current work, we will take into consideration the relationship between thought and the world when needed and we will not present a full-fledge account of the issue.

The possibility of maintaining the distinction between language and thought and to understand how the two diverge even if endorsing a cognitive view of language (but without falling completely into the "cognitive-priority" view of language) come for linguistics, and, we will argue, especially from the Saussurian tradition. Within the Saussurian framework, language is not just a cognitive phenomenon, but it essentially arises from the standardization of language uses in a community of speakers. Hence, language is inherently a social and interactive phenomenon. The structure of language, the *langue*, emerges as an abstract system precisely from the repetition of standardized uses. Language is both internal to our mind as well as external, essentially residing among the speakers of each community. As we will see in more details throughout this work, the Saussurian perspectives (interpreted in the light of De Mauro's readings) allows us to think about a separation between linguistic and cognitive representations (and hence about a language-specific level of linguistic meaning) without incurring into a modular view of the mind and in line with current approaches in cognitive linguistics (as, for example, with usage-based theories of language). Language as a system of meaning is neither completely separated nor completely reducible to domain-general processes and to non-linguistic knowledge.

The second question we outlined at the beginning of this section is concerned with the mechanism of the influence of language on thought and it is inherently dependent on how we define meaning in the first place (a task not often explicitly addressed in relativistic studies and approaches). We will here try to briefly outline the three main ways in which the mechanism of interaction has been formulated within the relativistic enterprise: the Whorfian *linguistic analogy* and *cognitive appropriation* mechanism, the neo-Whorfian *language as interference*, and the most recent *language as ingredient*.

Throughout his work, Whorf never really offered an explicit account of *how* (i.e. the mechanism by means of which) language should influence thought. The most renowned attempt to bring to light the implicit theory underlying Whorf's analysis is proposed by Lucy (1992). Let us briefly recall Whorf's perceptive on language and linguistic meaning; as we have already seen, Whorf endorsed a view of language as a system of interrelated patterns (*fashions of speaking)* that gives rise to language-specific conceptualizations and segmentations of what he refers to as the *flux of experience*. Hence, language is not identical to thought, but it rather guides people's habitual

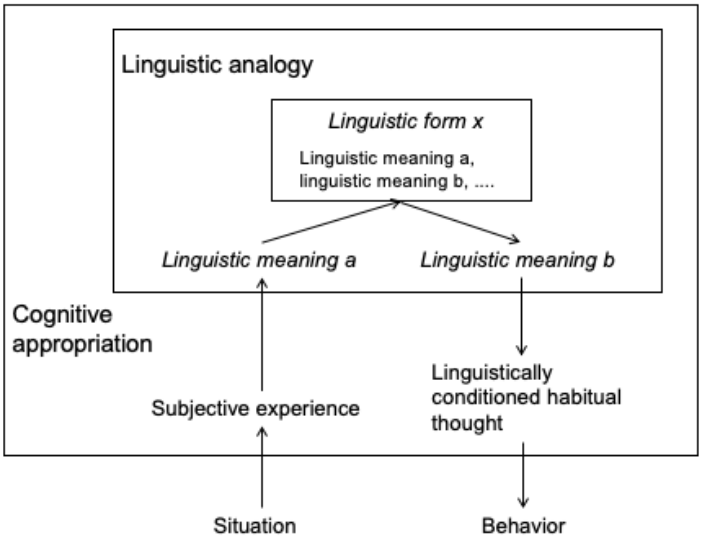


Figure 3.Whorf's "linguistic analogy" and "cognitive appropriation" mechanism as outlined by Lucy (1992; schematic adapted from p.47 Figure 1)

everyday thought[28] in their interpretation of experience. The mechanism of this influence is outlined by Lucy as follows (see figure 3): language unites demonstrably different aspects of reality by giving them similar linguistic treatment. Language operates a classification of reality based on *linguistic analogies*:

> [s]ome aspect of a "situation" is immediately and subjectively "experienced" by a speaker of a particular language. This experience can be interpreted in terms of a 'linguistic meaning a' which corresponds to a *linguistic form x*. The *linguistic form x* also corresponds to other linguistic meanings such as 'linguistic meaning b.' By virtue of 'a' and 'b' being grouped together under x, they, and the experiences which they label or refer to, are seen to be analogically "the same" even though in some respects they are not in fact the same at all (Lucy, 1992, p. 47).

The importance of such linguistic classification is that the meaning of elements that are grouped together influence each other by means of being interpreted as "the same". Such linguistic

[28] One clarification is here in order. When Whorf analyzes the effects of language on thought, he is mostly referring to cultural patterns of habitual beliefs and behaviors. In contrast, with the rise of the psycholinguistic approaches Whorf's perspective on thought was highly criticized and thought understood as essentially an individual phenomenon. The primary goal of psychologists is to demonstrate the validity of the proposed linguistic effects on the thought and behavior of individuals. Hence, since Lenneberg's reformulation of the relativistic hypothesis, the emphasis is on individual processing potentials, and not on the share content of interpretative categories of thought which had been the focus of anthropological linguists (for a discussion cf. Lucy, 1992).

analogies are used in thought as guides in the interpretation of and behavioral responses to experienced reality: the *cognitive appropriation of linguistic analogies* (i.e. the influence of language on thought) happens when the speaker uses these linguistic classification in interpreting or reflecting on his or her immediate subjective experience, and hence the meaning of 'a' is called into play when responding to experiences properly associated only with 'b', since 'a' and 'b' are analogically interpreted as "the same". For instance, this is the mechanism that underlies the famous case of the word "empty" in the fire-causing example:

> Yet the "empty" drums are perhaps the more dangerous [in comparison with the full drums], since they contain explosive vapor. Physically the situation is hazardous, but the linguistic analysis according to regular analogy must employ the word 'empty,' which inevitably suggests lack of hazard. The word 'empty' is used in two linguistic patterns: (1) as a virtual synonym for 'null and void, negative, inert,' (2) applied in analysis of physical situations without regard to, e.g., vapor, liquid vestiges, or stray rubbish, in the container. The situation is named in one pattern (2) and the name is then "acted out" or "lived up to" in another (1), this being a general formula for the linguistic conditioning of behavior into hazardous forms (Whorf, 1956, p. 135).

As Lucy points out, Whorf's perspective yields a vision of language as a formal classificatory device that is both distinctively linguistic as well as internally interacting with other aspects of cognition. Unfortunately, Whorf never addressed the issue of when and why such appropriation of language structures occurs and what might be the place and format of linguistic classifications deriving from the mechanism of linguistic analogy in cognition.

This last aspect became fundamental in psycholinguistic approaches that, as we have previously mentioned, view language mostly as a cognitive phenomenon (even if independent from thought). The standard view of neo-Whorfian approaches is that language and thought provide two separate set of representations. Starting from the work of the two psycholinguists Brown and Lenneberg, language was essentially reduced to the denotational value of lexical items, that is, the array of objective stimuli they could refer to. Eventually, this denotational emphasis led to the complete elimination of any concern with structures of meaning or grammatical differences among languages. As an example of this reduction, in Lenneberg and Brown's account, the only linguistic variable was "code efficiency": "they claimed that inter-subjective agreement in the denotational application of individual lexical items can provide quantitative index of accuracy in perceptually distinguishing among the pertinent objects of reference" (Lucy, 1992, p. 216). Even

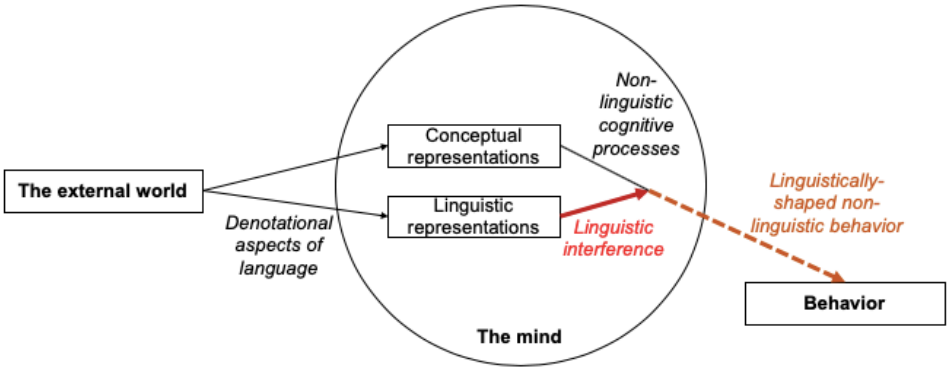


Figure 4. Schematic representation of the neo-Whorfian (general) mechanism of the influence of language on thought.

after cross-linguistic studies and structural aspects of language were regained in approaching the linguistic relativity hypothesis, the mechanism envisaged by Brown and Lenneberg remained substantially unaltered. Linguistic representations, that are thought of in terms of their denotational applications, interfere with otherwise non-linguistic cognitive processes in giving rise to linguistically-shaped non-linguistic individual behavior (see picture 4 for a schematics). From this general schema, all the subsequent work in linguistic relativity aroused. Clearly, this is just a very general interpretation of the basic mechanisms that underlies neo-Whorfian approaches, but, as we have seen, particular instances of how this interference might actually happened can be of various, often very diverse, sorts.[29] For example, if in the color domain we are mostly focused on codability and on the effects of linguistic categorization on perceptual processes (which are taken to be intrinsically non-linguistic), the case of space representation might be slightly different. Suppose someone speaks a language with default geocentric spatial reference. It follows that whenever he or she is observing a spatial relationship between some entities, he or she must pay some attention to the geocentric configuration of that relationship. More generally, linguistic requirements cause speakers of geocentric languages to mentally encode certain aspects of spatial relationships so that those relationships can be subsequently described in a fluent manner. Such relationships will inevitably become relatively salient for them, when contrasted to speakers of languages that do not rely on geocentric language to the same degree. Hence, speakers develop certain mental habits

[29] We have reviewed various kinds of relativistic effects following Wolff and Holmes (2011) in the previous section.

(habitual patterns of focus) because of the exigencies of one's native language that could then foster default patterns of conceptualization.

While there are a variety of proposals on the sorts of ways in which linguistic dissimilarities may gradually result in differences of thought, it is fair to say that more work needs to be done to clearly establish the processes through which relativistic effects come into existence. This is true, at least in part, because the way in which such effects develop likely varies in accordance with the kinds of effects in question.

In most recent years, it has become clear that a straightforward distinction between language and thought, linguistic representations and conceptual representations, is not easily obtainable especially in light of recent accounts of meaning in cognition as essentially encyclopedic in nature. Cognitive linguists have often shifted towards a different idea that we have already targeted as the "cognitive-priority" hypothesis: concepts (however defined) come first and languages labels (sub-portions) of them. As of consequence, finding a place for separated (to some extent independent) linguistic representation in cognition have become more and more complicated. Within a cognitive framework, a possible alternative to thinking about words deriving meanings by mapping onto a separate conceptual landscape, is to think of words as helping to construct meaning, a framework often glossed as words-as-cues[30] (Elman, 2004, 2009; Lupyan, 2016; Lupyan & Bergen, 2016; Rumelhart, 1979). On this view, the meaning of words is "revealed by the effects it has on [mental] states" (Elman, 2004, p. 301), i.e. on the same mental states that are activated in response to nonverbal stimuli: sensory states, motor states, affective states, and the combinations thereof. Hence, words become part of a much wider and intricated process of doing-meaning that involves a range of bodily, linguistic, cultural, social factors, and so on, all contributing to (being ingredients of) our mental contents. As we will discuss in the next chapters, the format, place, and processes involved in mental representation (linguistic or not) is one of the issues in the definition of what has been termed the "mental lexicon". In the word-as-cues hypothesis, word meanings do not necessarily reside in our head, but are rather thought of stimuli that operate directly on mental states. The role of language in shaping our concepts can be then re-addressed as follows: what do words (and experience with language more broadly) do to

[30] We will come back to this notion of word-as-cues in the next chapters and especially to Elman's perspective in chapter 4.

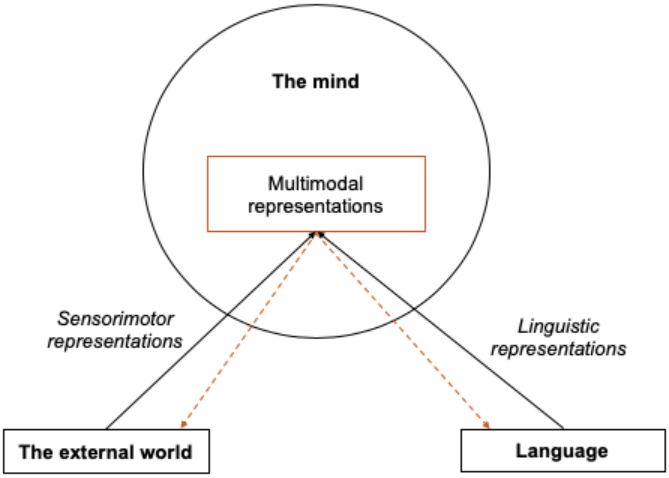


Figure 5. Schematic for language-as-ingredient perspective. Multimodal representations comprise elements from sensorimotor and linguistic experiences (as well as more broadly social, cultural, emotional experiences, and so on – that we not representing in this schematics). The full-lines arrows indicate the bottom-up influence of the various sources of experience in giving rise to our mental content. The dashed-line arrows indicate the top-down effects of multimodal representations in our interaction with the world. Hence, the influence of language goes both ways: both in the way in which we build our knowledge from experience and in the projection of this previous (linguistic as well as non-linguistic) knowledge onto our ongoing experience. The schematics is of course extremely simplified given the complexity and variety of the processes involved, but it just has the role to give a general overview.

our mental states that comprise semantic knowledge in the course of leaning and using language? As we have seen in the previous paragraphs, Lupyan's proposal of word labels as augmenting and enhancing categorical representations, and abstraction more broadly, can be taken to fall under this specific instance of the question of the relation between language and thought.

On a similar line of research, a growing body of studies in psycholinguistics and psychology has been jeopardizing the assumption for which conceptual representations and linguistic representations are distinct and separated, by suggesting that conceptual knowledge depends both on embodied and situated groundings as well as on information statistically learned from linguistic experiences (Andrews, Frank, & Vigliocco, 2014; Andrews, Vigliocco, & Vinson, 2009; Boroditsky & Prinz, 2008; Louwerse & Jeuniaux, 2008; Vigliocco, Meteyard, Andrews, & Kousta, 2009). Representations are multimodal in the sense that they not only arise from the integration of various sensorimotor experiences, but also in that they are shaped by the language we use (see figure 5). Thus, research revealing that linguistic experience essentially is one of the ingredients that contributes to the acquisition and construction of our concepts forces a

reconsideration of both the supposed distinction between language and thought as well as of the role of language in shaping not as much our non-linguistic cognitive processes, but rather the very content of our mental representations, which, in turn, are then called into play each and every time we interface with the world, giving rise to top-down effects both in perception as well as in other supposedly non-linguistic cognitive abilities. The word-as-cues proposal and the language-as-an-ingredient cannot be taken as completely overlapping or as suggesting the exact same mechanism of influence. We have presented them together here because they both call into question the possibility of thinking of language in cognition as a separate set of representations: in the first, words are taken to activate and operate on mental states; in the latter, language is taken to contribute to the formation of mental states in the first place. The mechanisms by means of which words operate as cues or by means of which we learn from experience with language, and what we actually learn, are still largely unexplored and more research is needed in order to address these ongoing issues.

The aim of our review so far is not to express a preference on which one of the mechanisms or of the various proposal is more likely to be correct in addressing what is going on at a cognitive level when we use language or when we create meaning. In the next final section, we will try to draw the structure of our main argument and of the themes that will be addresses in the next chapters. We will summarize the main issues that arose in this introductory chapter, outlining how they take us all the way through an in-depth discussion of the possibility of identifying a "language-specific" level of linguistic meaning, which will be the base for a range of possible hypothesis on the role played by language in shaping and structuring our mental representations.

1.3. An outline on the structure of our thesis

In this introductory chapter, we have touched on many different points and raised many issues that we regard as still central in the study of linguistic relativity as well as in the study of the relation between language and thought more generally. We will here briefly summarize the core aspects of our discussion so far and reveal how they will unfold in the rest of this work.

We opened our discussion with the linguistic relativity principle/hypothesis as our main topic, we progressively shifted away from it, and we will almost completely leave it out of our discussion from here on. There are multiple reasons for this shift in perspective: as we have argued, the linguistic relativity hypothesis should be thought of not as a unitary phenomenon, but rather as a family of approaches all sharing a very general stand on the issue of the relationship between language and thought: language does not merely allow us to convey our thoughts, but it also contributes in shaping our mental content and/or our cognitive abilities. More generally, approaches that fall under the general umbrella of linguistic relativity are interested in evaluating the possibility of a cognitive function of language. How and by means of which mechanisms this influence of language of thought might get actualized highly depends on the definition of language and thought we assume. Hence, it is clear that every approach in linguistic relativity necessarily rests on specific assumptions about the nature of language and of cognition. Within the linguistic relativity debate, however, little attention has been devoted in building an explicit theoretical framework in which both language and thought could be accurately discussed.

Especially on the side of language, linguistic relativity approaches have generally used an extremely simplified notion of linguistic meaning: although assuming a perspective radically different from the one generally endorsed within the cognitive science, i.e. that language simply maps onto pre-organized concepts, they have been mostly interested in the referential/denotational aspects of words. The separation between language and thought (that we inherently need to assume if we want to talk about an influence of the first on the latter) is generally grounded on the evaluation of how languages encode specific aspects of the reality. We have triec to show how the focus on the referential function of language does not efficiently allow us to establish a distinction between language and thought, especially if we take into consideration recent advances in the study of cognition and of meaning in cognition. Hence, the main question we will be addressing

in our work is concerned with the very possibility of establishing a distinction between language and thought on solid theoretical grounds. Assuming that language and thought are actually distinct but not completely independent (given that language is inherently cognitive and cognition hardly extricable from language), where should we look to establish where the two differ and where and to what extent they influence each other? We will argue that focusing on non-strictly referential aspects of meaning in language, namely on the structural, systematic (in a Saussurian sense) aspects of language, may provide interesting insights into the possibility of identifying a "language-specific" level of meaning and, consequently, into the study of the relation between language and thought.

Furthermore, as our work will hopefully show, the individuation of a specifically linguistic level of meaning would take us in the direction of evaluating what aspects of language might shape our mental representations and cognitive processes from the "bottom-up", i.e. language and language structure seems to be implicated in the way in which we learn and construct our mental representation, becoming one of the ingredients that gives form and content to our interaction with the environment. In order for this perspective to hold, language needs to be thought of as something not only internal to our mind, but as well as something that is non reducible to our private mental representations, as something we find outside of us, in the world we are embedded in, and from which we learn not just how to use language for communicating our thought, but also to shape our knowledge and interaction with the world.

Hence, the structural aspects of language shared by a community of speakers become fundamental to understand how we construct our view of the world. We will argue that computational models arising from a distributional account of linguistic meaning (Boleda, 2020; Lenci, 2018), originating from the idea that "we shall know a word by the company it keeps" (Firth, 1957), can give us a strong theoretical as well as methodological perspective on how to address the structural aspects of language starting from a usage-based account of meaning. In very general terms, distributional models of semantics attempt to discover structure in language by tracking the contexts in which words are used. From this general insight, on the side of linguistics, distributional semantics allows us to extract representations of word meanings from large text of corpora and, on the cognitivist side, to formulate specific hypothesis on the format of semantic representation, on the way they are built and processed by the human mind. The idea of language-as-ingredient, enriched with a relational account of the language-specific level of linguistic

meaning, addressed by means of distributional semantic approaches, will constitute the main theoretical argument of our present work, on which we will build to integrate a systemic view of language in the methodological approaches to the study of language in cognition. Our argument will be structured as follows.

Chapter 2

In chapter 2, we will undertake the issue from the very beginning, but without focusing as much on the relativistic perspective in and for itself. Firstly, we will re-frame the question at the interface between the world, cognition, and language, showing how the three dimensions are inextricably connected and inherently involved in the study one of the other. Secondly, we will devote particular attention to some fundamental terminological issues: we will discuss our operative notion of meaning, setting some boundaries in what we mean by *meaning* and *semantics*, and we will review some of the most outstanding definition of *concepts* in relation to *linguistic meaning*. This terminological (which is indeed at the same time theoretical) detour is fundamental given the centrality of these notions in many different disciplines, and since the definitions given and the uses made of these notions can be very different across fields of research. Especially within our perspective, that sets itself at the intersection of linguistics and cognitive sciences, it is particularly important to clear how we will be using these terms in regard to and in contrast to other possible uses. We will then turn to the first part of our analysis, which will take us to the study of the structure of language; we will engage in the study of the relation between language and reality and we will reveal how a systemic and relational perspective on meaning can guide us in understanding how language carves nature at its joints. In conclusion, following De Saussure, and De Mauro's readings of the Saussurian perspective, we will argue that, from a linguistic perspective, the relationship between language and reality should be understood as a constantly dynamic process of meaning creation, that we will gloss as the process of "doing meaning". The main goal of this chapter is to set out the theoretical outline for the proposal that will be developed in the next chapters, assuming a linguistic perspective grounded in a systemic, relational, and dynamic view of linguistic meaning. The main outcome of this chapter will be that of recognizing the centrality of the study of the structural aspects of language (and especially of the lexicon) in the study of the relation between language and thought, also tangentially allowing us to restrict our field of inquiry to the study of the structure of language.

Chapter 3

In chapter 3, we will assume a historical perspective on the study of the structure of the lexicon and of meaning more broadly. The main goal of the chapter is to review how different approaches have been addressing the notion of meaning and of the structure of meaning in language. In the first place, we will start off with the Saussurian notion of system as associative (paradigmatic) and syntagmatic relationships. In the second place, we will show how the notion of semantic (lexical) field, arising out of the Saussurian paradigm, essentially reshaped the study of the structure of language leading towards a holistic (and locally holistic) view of meaning both in the theory and in the methods of analysis of the lexicon. In the third place, we will turn to the compositional approach and to the integration of lexical fields and compositional approaches in the study of meaning. These last two approaches have been fundamental in the relational study of linguistic meaning inasmuch as they provided a way to establish a methodological (as well as theoretical) approach to identifying the structural aspects of language. Subsequently, we will follow the historical development of studies in lexical semantics and we will briefly review the cognitive twist, analyzing the first attempts to integrate componential analyses and cognitive semantics. Afterwards, and in conclusion, we will review the changes that happened in the study of meaning after the second cognitive revolution which re-installed linguistic semantics as a central topic in cognitive linguistics, but essentially equating it with the study of concepts and cognition. With the first and the second cognitive revolution, language (and especially linguistic meaning) loses its specificity and the language structure becomes essentially a window onto the conceptual structure.

Chapter 4

One of the most outstanding outcomes of the cognitive revolution(s) is that of viewing language as a cognitive phenomenon. In the third chapter, we will discuss some of the possibilities arisen within cognitive approaches of identifying and evaluating the structure of language in cognition. Viewing language as a cognitive phenomenon led to one very general need: it seems clear that for language to work in cognition, i.e. for us to process language in both production and comprehension, we need to be equipped with a storage of linguistic information that can be easily retrieved each time language is activated. This repository of language in cognition is generally called the *mental lexicon*. In this chapter, we will review some of the main accounts that have been proposed of the mental lexicon for the purposes of language processing in cognition. Hence, in

this chapter, we are interested in the nature and structure of the lexicon as much as in how language is processed at a cognitive level. We will discuss the metaphor of the mental lexicon as a dictionary within the generative and modular framework and, subsequently, we will outline how this metaphor of the dictionary had to evolve in the light of renewed considerations on the nature of both language and thought. Especially the shift from serial models of language processing to interactive models of language processing will open up the way to a change in the metaphor used to describe the mental lexicon: language is not stored and operates as a dictionary, but rather as an interconnected network. We will present and discuss local and distributed network models of language processing and storage and we will conclude by evaluating one of the possible outcomes of this view of language as a network; i.e. that we might not actually need a mental lexicon at all, a storage of any sort, in order to process language inasmuch as words are thought of as not having meaning, but rather as being cues to meaning. In the final section of the chapter, we will leave aside how language might be processed in the mind and focus on the kind of mental storage we need if we assume an encyclopedic perspective of meaning (which is at the heart of the second cognitive revolution). We will present the case of free word associations and semantic networks based on association, trying to discuss what kind organization they do in fact reveal of meaning in cognition. In conclusion, this chapter is mostly concerned with meaning and its structure from a cognitive perspective, in which meaning is defined as inherently internal to the mind and its structure dependent upon the needs and limitations ingrained in cognitive processes.

Chapter 5

In chapter 5, we will start off by arguing that a perspective of meaning as strictly internal to the mind does not allow us to grasp some fundamental aspects of linguistic meaning and, more to the point, it prevents us from establishing a possible level in which language and thought can be distinct. Firstly, we will briefly summarize positions in cognitive linguistics that advocate for meaning as conceptual and meaning as conceptualization, pointing out some of the shortcomings of these views. In the second place, we will introduce a perspective put forth by some approaches in cognitive linguistics, and especially in cognitive grammar, known as usage-based theory of language. We will try to show how usage-based perspectives open up the way to a reformulation of linguistic meaning within cognitive approaches that does not necessarily flatten out linguistic meaning onto conceptual meaning. Building of the insights provided by usage-based approaches,

we will introduce our proposal regarding the possibility of identifying a "language-specific" level of linguistic meaning in the structural, relational aspects of the language system; i.e. we will try to show how the view of meaning introduced in chapter 2 can give us one possible way of answering the question of the separation between language and thought. Based on this theoretical perspective, we will argue that distributional semantics offers us the methods and theoretical grounds to investigate the structure of meaning in language by extracting the relational structure of language from language uses. Our introduction to distributional semantics will focus on two main points: in the first place, we will discuss the historical roots of distributional semantic approaches revealing their inherent relationship to the structuralist and, more specifically, Saussurian framework, and we will point out where the two differ. In the second place, we will briefly present computational distributional models of linguistic meaning from a linguistic perspective: we will introduce some of the most recent advances in computational techniques and what aspects of the relational structure of linguistic meaning they allow us to address and represent.

Chapter 6

In chapter 6, we will discuss distributional semantics in cognitive research on meaning. We will present the "strong" distributional approaches, according to which distributional semantics allows us to formulate specific hypothesis on the format of semantic representation in cognition. On the one side, we will review accounts of how distributional representations of meaning perform against some of the standard cognitive benchmarks, hence allowing us to make appropriate predictions about meaning in cognition. On the other side, we will openly face the question about the role of distributional representations in re-shaping the question about the interaction of language and though: assuming that distributional representations allow us to represent the "language-specific" relational aspects of linguistic meaning, the question that immediately follows is how these representations might interact with cognition more broadly and in what sense they may be implicated in shaping what we know. The way in which we will reframe the question about the relation of language and thought, starting from the analysis of the language structure in language use, will be mostly concerned with investigating to what we owe what we know, i.e. where we get our knowledge from and to what extent may the "language-specific" level of linguistic meaning be involved in giving rise to our mental representations. In the first place, we will address some of the main limitations of distributional approaches if considered as comprehensive accounts of

meaning and in light of recent advances in the study of cognition. In the second place, we will turn to the second fundamental claim of our proposal: both language and cognition should be thought of as multimodal. Language is multimodal in the sense that it is both symbolic and grounded: it encodes perceptual-sensorimotor knowledge as well as it maintains a level of language specificity that cannot be reduced to the content we get from other modalities. Cognition is multimodal in the sense that our mental representations are derived from both sensorimotor and linguistic experiences. In conclusion, we will present one hypothesis stemming from the discussion we outlined: the structural aspects of language may be one of the main sources from which we learn from our experience with language. As a first possible way to test this hypothesis, i.e. that the structural aspects of language are one of the key factors in language from which we learn, we will present an empirical study of rapid word learning from morpho-syntactic and distributional cues. Although it does not allow us to draw any conclusion about the specific role of the language structure in language learning, this study suggests one possible way of empirically investigating the role of the "language-specific" level of linguistic meaning in giving rise to our mental representations (in this case, for specific instances of language). In conclusion, our proposal aims to set out a possible line of research on the relationship between language and thought by viewing language as an ingredient that, combined with many other aspects of our interaction with our internal as well as external world, gives rise to our cognitive representations and to the structure of our knowledge. It is clear that these are just the first steps into the line of inquiry we have been proposing. Further research is needed to assess the validity of our claims as well as to evaluate the testable hypotheses that arise from our perspective.

Chapter 2

Theoretical foundations for a study of language and cognition

---

2.1.Prelude: on language and thought

2.1.1. The position of the problem

In the *Hitchhiker's Guide to the Galaxy*, our planet was commissioned, paid for and run by mice, a protrusion in our dimension of vastly hyperintelligent pandimensional beings, interested in solving the problem about the meaning of life. Interestingly, the formulation of that wretched thing called the Ultimate Question could be obtained only through the observation of and the experimentation on human behavior and (eventually) the human brain. Imagine for a moment to be one of those mice that was sent to Earth to understand how and why humans behave and act; what would your first impression be? The complexity of human life is clearly an astonishing fact also from the perspective of an alien observer (why bother creating it this way and studying it for ten million years otherwise?); and not only in the obvious observation of the complexity of cultural and social environments in which humans live, but, at the same time, in the very constitution of their relationship with the world. From a layperson perspective, it might seem like we perceive, experience and talk about things as we actually encounter them in the world that surrounds us. But if we start taking a closer look at the nature of our experience, the relationship between the external reality and how humans perceive and categorized it does not seem that straightforward.

Think about color; our experience of color is far from obvious. How do we perceive it? Is color out there in the world or is it a projection of our interpretative cognitive, cultural and linguistic schema? Does color pertain to the realm of nature or to the one of culture? In the introductory chapter, we have already mention color perception and categorization as one of the most renown benchmarks of linguistic relativity. The study of color perception and categorization comprises a microcosm of research paradigms and theoretical issues spanning across a wide

variety of disciplines including cognitive sciences, psycholinguistics, neurosciences, physiology, philosophy of perception, philosophy of language, and the list could go on. Such a longstanding interest in the perception and categorization of color is due to the fact that color provides a paradigmatic domain in which science and human experience, nature and culture, mutually intersect. In fact, on the one side, color can be analyzed as it appears in the physical world; it can be described as a combination of six basic colors (red, green, yellow, blue, black and white) varying across the dimension of hue, saturation and brightness. On the other side, we mostly perceive color as an attribute of things in the world and there seems to be no simple one-to-one mapping between light flux of various wavelengths and the colors we perceive surfaces to have; i.e., surface reflectances are not to be found in some pre-given world that is independent of our perceptual and cognitive capacities (Varela, Thompson, & Rosch, 1991).

Furthermore, a recent trend of studies has revealed that perception itself is not an isolated process in our minds. We already mentioned the case of top-down effects of linguistic categories on color perception in the previous chapter. More broadly, it is now widely recognized and supported by empirical evidences that neural processes are structurally interactive (Foxe & Simpson, 2002; C. D. Gilbert & Sigman, 2007; Koivisto, Railo, Revonsuo, Vanni, & Salminen-Vaparanta, 2011; Reynolds & Chelazzi, 2004) and that the brain is only grossly hierarchical allowing lower level processes (as perception) to be influenced by later processing (Churchland et al., 1994). This roughly means that visual perception, including color perception, can be affected by non-perceptual influences such as previous knowledge about the world, goals, expectations and even linguistic knowledge (Lupyan & Clark, 2015). Taken together, these considerations suggest that we may not be able to explain what color is and how we perceive it if we locate it in a kind of experience of the world that is independent of our perceptual and cognitive capacities (Varela et al., 1991).

When generalizing this conclusion beyond the domain of color, at stake there is a more general understanding of cognition and of how humans organize their systems of knowledge. To this regards, we can distinguish between two general positions in the history of ideas that set themselves at the opposite extremes of the debate: on a one side, the strongest realistic accounts view cognition as the mere reconstruction of a pre-given and pre-organized external world (i.e. the organization of our knowledge is "out there" in the world); on the other side, according to more idealistic perspectives, cognition projects its own structure on the world and reality becomes a

mere reflection of our internal cognitive structures (i.e. the organization of our knowledge is "in here" in our minds). At the current state of the art of the studies on the cognitive architecture as well as on perception and other cognitive functions, a clear-cut distinction between what is "out there", in the external world, and what is "in here", in our minds, is hardly sustainable. The structure and content of our knowledge depends both on our biological marks as well as on our perceptual and cognitive capacities[31].

And yet this is not all there is about color. Colors are categories and, most importantly to us, they are linguistic categories. Even when we think we might have found universality in the perception of colors, cultural and linguistic differences come out as a striking evidence that cannot be ignored if we aim to understand how we conceptualize, and probably even perceive, colors (for a review of studies cf. V. Evans, 2014).

On the one side, a general question addresses the fact that we organize experience into categories and give names to these categories. This gives rise to long-standing controversy of whether the linguistic categorization of reality (as for example in our illustrative case about the color domain) is independent (or to what extent it can be considered independent) from the organization of the external reality. Linguistic variation in the segmentation of the experiential and conceptual space of color has been observed across different cultures and over many centuries of research. From the studies on the "wine-dark" or "wine-looking" sea in Homer's *Iliad* and *Odyssey* (Gladstone, 1858), moving on to Hjelmslev's study on the segmentation of the chromatic field in English and Welsh (Hjelmslev, 1961), to the research on more exotic cultures as the two-color-terms system of the Dani tribe of New Guinea ([Rosch] Heider, 1972) and to the simple observation that English has only one linguistic category for blue whereas Italian speakers separate light-blue (*azzurro*) from dark-blue (*blu*) and Russian speakers do so to (*siniy* for dark-blue and *goluboy* for light-blue), it seems that languages have significant freedom in the way in which they categorize reality. Nonetheless, this is not a freedom without constrains; although considerable variation over color category boundaries is observed across cultures, the similarity among languages in the choice of foci are still far too striking to be simply dismissed (Berlin & Kay, 1969; Regier, Kay, & Cook, 2005). Even more, as far as we know, there is no language that has a category

[31] In most recent years, a possible third view has been proposed that set itself programmatically in between these two extremes. The theoretical and research paradigms that go under the general label of "embodied" theories of cognition are an example of this fruitful third view (for a useful discussion see: Varela, Thompson, & Rosch, 1991).

that includes two areas of the field (like yellow and blue) but that would exclude one area in between the two (for example, green) (Roberson, 2005). These observations suggest that color categories are not simply to be found in a pre-organized world, but they arise from a combination of biological constrains and linguistic (but also more broadly cultural) structures.

On the other side, as we have already discussed in chapter 1, a now well-known trend of studies has devoted its attention to whether and how the linguistic categorization of the color domain shapes our perception of colors or is involved in other cognitive functions (for a more complete overview on the history of the debate on colors and linguistic relativity see Deutscher, 2010). It has been variously argued that the language we speak provides us with lens through which we perceive and categorize reality. In the recently-born subfield of psychology generally referred to as psycholinguistics, the relationship between language and thought is of crucial interest. In the previous chapter we already discussed some of the main concerns and issues regarding the study of the influence of language on non-linguistic cognitive processes; let us recall some interesting findings regarding the influence of linguistic color categories on color perception. Paul Kay and Willett Kempton (1984) showed how speakers of languages that treat blue and green as different colors (like English) are biased in perceiving shades near the blue-green border; speakers tend to exaggerate objective distances across the border and to underestimate distances on the same side of the border. Reaction times in recognizing the difference between certain colors have shown that there are measurable differences between the ways in which the visual systems of Russian and English speakers react to shades of blue (recall that Russian speakers differentiate linguistically between dark and light blue whereas English speakers do not) (Winawer et al., 2007). Finally, linguistic enhancement of color contrasts is one of the forces that determine what we consciously perceive (Maier & Abdel Rahman, 2018). Even though these studies do not settle once for all the long-standing issue on the influence of language on cognition, they convincingly show that the categories acquired with language play a significant role in the way in which we perceive and understand the world, prospecting a role for language that, most probably, goes beyond allowing us to talk about the external world.

There is something more that can be pinpointed about language and its role in constructing our system of knowledge. When we think about a color or when we imagine it, there is more to it that what is actually visible; every object is part of a weave of relationships that are not in the object itself but rather depend on our general knowledge about the world and on our knowledge of

language (Liuzza, Cimatti, & Borghi, 2010, p. 99). To what do we owe our knowledge of color? How do we learn the meaning of color words? How do we know that two colors are more similar compared to a third? How do we know that the sky is blue, and the grass is green? Evidently, many of us know all these from our perceptual experience of the world. But individuals who are congenitally blind also know the characteristic colors of many objects as well as the relationships between colors, suggesting that linguistic experience is a powerful source of information even for those domains that we generally considered "property" of sensory and motor experience (Lupyan & Lewis, 2017).

In the futuristic scenario described above of alien-mice studying our species, they could learn a lot about how we perceive and experience colors by looking at how we organize the color domain linguistically and by hearing us talking about it (note that a fascinating peculiarity of those mice is that they understand and speak English). What would then be the judgment of the alien-mouse trying to understand human cognition and behavior after this brief dive into how we perceive and talk about color? It seems fair to conclude that if we want to understand the kind of relationship humans entail with the world, we cannot avoid considering the role played by language. One of the aspects that seems most relevant about language in the framework we outlined above is how language carves up the external reality, its constrained freedom in selecting categories and categorical boundaries and how linguistic meaning is constructed and organized in relation with, and in comparison to, the structure of the external world. As we already mentioned in the introduction and we will better show in the next sections, language not only reflects the joints of nature, but *carves joints into nature*. Our aim in this chapter is to understand how and by means of which internal mechanisms languages might do that (from the perspective of linguistics).

If, on the one side, languages can be analyzed in the way in which they carve up nature, on the other side, there is definitely more to it. Language is, at the same time, an aspect and a product of cognition; i.e. language cannot be investigated independently from cognition more broadly as it is inherently a cognitive phenomenon. The semantic of natural languages is probably the central point of this interaction: in the process of signification, in the way in which humans "do" meaning, we can glimpse at this relationship between language as a product, i.e. as a stable system of meanings shared among speakers of the same linguistic community, and language as an aspect of that ensemble of cognitive mechanisms that constitutes and drives our actions and behaviors. Throughout our work, we will mostly be focusing on the meaning of words. The meaning of words

in neither only "in the head" of the speaker in the form of a mental lexicon nor only "out there" in the world of spoken and written humans' interactions; linguistic meaning seems to arise from the combination and interaction of both spheres (Basile, 2001). Language is a constantly dynamic process constitutively involved in shaping our mental as well as socially shared contents in a continuous negotiation with the joints imposed by nature. This is our starting point as well as the general assumption that underlies our work. To understand the interface between language and thought, and what language may do for us at a cognitive level, we will have to delineate some of the fundamental aspects of the relationship between language and the external world, although without losing touch with its intimate connection with the realm of thoughts and concepts. As we will better discuss below, language does not simply map onto a pre-organized, pre-given external reality or internal mental concepts; within the constrains given by the world and by our perceptual and cognitive devices, historical-natural languages segment the space of the content (all that can be thought of or talked about) in different and often non-overlapping ways. We mentioned the linguistic variability that can be observed in the domain of color; even if we cannot think of the color space as a completely unstructured *continuum* (we have seen how few categories and color foci could be identified as universal across languages and cultures), it is nonetheless true that it is only loosely segmented and constrained and even where those constrains seem to be in place, languages can carve up nature otherwise (recall the example of the two-color-terms system of the Dani of New Guinea). Whether or not these categories have an influence on perception or cognition more broadly is a different question, which we will try to deal with later in our work. Our aim in this chapter is to travel through those linguistic theories that consider language as more than a mere mapping device and to outline the conception of linguistic meaning that arises from them. This would lead to a discussion of some fundamental notions that, in turn, will programmatically shape the rest of our argument: namely, meaning and semantics. Before turning to this main goal, we ought to pause for a second and try to set some guidelines for the inquiry on linguistic meaning and its relation to cognition.

2.1.2. Concepts and words

As we have pointed out multiple times already, one of the fundamental issues in the study of language and cognition (and language *in* cognition) is the definition of the terms. This is especially relevant for the study of linguistic relativity, but it is an essential point of all inquiries into this relationship. In particular, the definition of the units of thoughts and of the units of language, and of the relationship between the two, is most probably the key theoretical crux of all studies at the intersection of language and cognition. Here, we will outline some of the fundamental theoretical positions on the notion of concept and on the relationship between concepts and language (and, especially, the lexicon of language). As it might be already clear, we are starting off by discussing relevant positions on one of the main themes the underlies our work, which is, once more, how we can disentangle language from thought in the first place.

Disputes about the nature of concepts (the kind of things concepts are, how concepts are structured, and where they get their content from) often reflect deeply opposing approaches to the study of the mind and of language. For obvious limitations due to the nature of our project, we will not present an extensive review of the different positions on the nature of concepts, which has been and still remains one of the most debated issue in philosophy and in the cognitive sciences (for in-depth reviews of the history of the debate we refer to Laurence & Margolis, 1999; Margolis & Laurence, 2015, 2019; G. L. Murphy, 2002; Prinz, 2002). This section will not provide an introduction to the study of concepts. This is an enormous topic that has a huge literature in philosophy, psychology, and the cognitive sciences, and one single chapter cannot even mention many of the relevant topics, much less develop them in detail. The goal of this section is to briefly introduce how concepts have been related to language (and more specifically to words) in psychological and cognitivist views. A propaedeutic introduction on the debate on concepts is obviously needed to pursue the goal of this session, but it is not our main focus. However, various theories of concepts will be addressed in the course of the chapters, as they will become relevant for the study of the structure of language *and/in* the mind; especially in chapter 3, we will see how various approaches within the cognitive sciences have been theorizing and defining concepts (and mental content more broadly) in different ways, from the *classical definitional view* to more recent account of concepts in terms of *prototypes* and *exemplars*, all the way to *embodied theories.*

What allows us to postpone the discussion about these different views of concepts to a later chapter and still discuss a definition of concepts and words? Psychologists have traditionally equated knowing the meaning of a word with the concept labeled by a word (Komatsu, 1992). In a nutshell, we want to meet the fundamental challenge of this work head-on: what is of interest to us is how the various definitions of concepts have been assuming, rejecting, or ignoring an underlying theory of language, i.e. how the relationship between language and thought has been negotiated in various accounts of the units of thought.

Any account of the human mind has to grapple, ultimately, with the nature of concepts. In the words of the philosopher and cognitivist scientist Jesse Prinz, concepts are "the basic timber of our mental lives" . Concepts, although fundamental for any study of cognition, as they are taken to be the constituents of thought, cannot be directly observed and, therefore, escape any direct experimental investigation. Experimental results do not directly indicate anything about concepts: it is always difficult do decide whether a particular observation is a function of the information provided by a given concepts, of the structure, or of the form of that information or of whatever process may be operating on that concept (Komatsu, 1992). Furthermore, it has been recently pointed out that in many experimental task, when we think we are tapping onto the content and structure of our conceptual knowledge, what we are actually investigating is people's knowledge associated with words (and not general-purpose concepts themselves) (Malt et al., 2015). Despite these difficulties in experimentally evaluating what concepts are and what their content is, it is generally recognized by psychologist and cognitive scientists that concepts are so fundamental for us that the whole cognitivist enterprise could be seen as a theory of concepts. For instance, "we negotiate our environments by thinking about them" and that "thinking itself subsumes such abilities as planning, reasoning, problem solving, deciding, and recalling" (Prinz, 2002, p. 2), which are taken to involve concepts as their basic constituents. Experimental results should allow us to specify at a minimum what kind of information must be available to people about the instances of a concept.

Taken together, these observational assumptions suggest that a theory of concepts and of what they are made of is needed if we want to have an adequate theory of our cognitive abilities. The MIT encyclopedia of cognitive sciences (1999) defines concepts as follows: concepts are

> the elements from which propositional thought is constructed, thus providing a means of understanding the world, concepts are used to interpret our current experience by

> classifying it as being of a particular kind, and hence relating it to prior knowledge. (p.176)

Leaving aside the issue of allowing propositional thought, the basic trait of concepts is that they can be thought of as a sort of "mental glue" (G. L. Murphy, 2002) that ties our past experiences to our present interactions with the world by providing us with the relevant categories that allow us to interact with our environment. Hence, theories of concepts go hand-in-hand with theories of categorization. In Prinz words, categorization "involves concepts: categorizing something is placing it under a concept or characterizing it by means of concepts" (2002, p. 12). Much of the experimental effort in providing an understanding of concepts has been based on studies of how we create and make use of categories: category identification is manifested when a person identifies the category under which an object belongs; category production is manifested when a person identifies which attributes an object possesses if it is a member of a given category (or, at least this is the operational definition provided by the MIT encyclopedia of the cognitive sciences). Psychologists believe that a theory of concepts should explain these abilities in a way that is consistent with empirical findings.

The majority of the theories in cognitive sciences maintains that concepts are psychological entities, namely mental representations. There are of course exceptions (Dummett, 1993; Peacocke, 1992; even key terms as representation, significance, and relation to the worlds are open to radically different interpretations: cf. Gallagher, 2008). As Margolis and Laurence (2019) argue, despite the terminological issue of which things ought to be granted the label "concepts", the same set of issues seem to arise. Even within theories that endorse a view of concepts as mental representations, the extent to which different theories agree with each other ends grossly here. Theories of concepts differ on the source, the format, the structure, and the content of mental representations. One captivating way of summarizing the different proposals that have been put forward in the study of concepts is that of Komatsu (1992, p. 550). We will prose a slightly different version of it here: the main question that theories of concepts have been addressing is what information is represented by, for example, the concept "bed" so that people are able to reason about beds, recognize instances of beds, and understand complex concepts such as "the pillow is on the bed".

The classical view holds that the concept "bed" represents information about the necessary and sufficient attributes of beds. The concept has a definitional structure ("the definition _is_ the

concept": G. L. Murphy, 2002, p. 15), it is composed of simpler concepts that express necessary and sufficient conditions for falling under "bed" (e.g. Katz, 1972; Katz & Fodor, 1963).[32] Consequently, all things that meet the definition are perfectly good members of the concept, and all things that do not fit the definition are equally "bad" non-members of the concept. In short, we should be equipped with a sort of check-list of attributes (i.e. of basic concepts that identify each concepts) against which we can check whether a certain entity is a member of it or not.

The classical theory has come under considerable pressure as it has outstanding difficulties in explaining a robust set of empirical findings (for reviews cf. G. L. Murphy, 2002; Smith & Medin, 1981). A battery of inventive experiments led cognitive and developmental psychologist to conclude that definitions are not psychologically real. Most of this evidence comes from studies of categorization, where it was found that people often group objects together on the basis of similarity judgements. Prinz terms theories of concepts that arose from these new finding "similarity-based theories" (2002)[33], which comprise *prototype theory* and *exemplar theory*. Very coarsely, according to the prototype (and family resemblance) theories (e.g. Rosch & Mervis, 1975), the concept "bed" is a summary representation that abstracts across specific instances of beds to give information about what beds, on average, are like. According to the exemplar view (e.g. Medin & Schaffer, 1978), the concept "bed" consists of representations of past exemplars of beds that a person has experienced, rather than a single summary representation of all beds.

Starting from the 1980s, psychologists began to question the role of similarity as the only (or even main) source of concept creation and to develop new theories of categorization. According to these theories, conceptual tasks bring to bear much more knowledge than previously thought. Concepts must incorporate that knowledge. In the "theory-based" model (e.g. G. L. Murphy & Medin, 1985; for other explanation-based theories of concepts cf. Lakoff, 1990), concepts are embedded in our broader understanding of the world. The theory-based representation represents knowledge about the relation of each attribute of a given concepts to others in a complex network of causal and explanatory links, represented in a structured frame or schema. The concept "bed" includes information about the interaction among beds, people, and other objects, as well as information about the (often casual) relationships that hold among the different properties of beds.

[32] One outcome of this view is the *informational atomism* proposed by Fodor (1990,1994), according to which almost all concepts are unstructured symbols that obtain their identity, in part, by carrying information about aspects of the environment (for a review Prinz, 2002).
[33] But see also Medin (1989), Medin & Wattenmaker (1987), Murphy & Medin (1985).

This brief review does not exhaust the variety of proposals in the theory of concepts, but, at least, it should give us a general overview on the complexity of the field.

A further problem is related to the format and source of our knowledge. A first view sees concepts as abstract, disembodied symbols. The representational format of concepts is qualitatively different from the sensory experiences concepts relate to, i.e. they are ultimately abstracted from the brain states that give rise to them, with information encoded in a different format (J. A. Fodor, 1975, 1998; Jackendoff, 1983, 1987; Pinker, 1984; Pylyshyn, 1984). A recent (reinvented empiricists; for a discussion cf. Prinz, 2002) account of concepts and conceptual structure emphasizes the psychological relations between the conceptual system and perceptual and motor states, giving rise to *the embodied mind thesis* that "minds profoundly reflect the bodies in which they are contained" (Shapiro, 2004; p. 167). Attempts to blur the boundary between perceptions and cognition can be found in the work of both psychologists (e.g. Barsalou, 1993; Barsalou & Prinz, 1997; Chemero, 2009; Gallese & Lakoff, 2005; Goldstone & Barsalou, 1998; Prinz, 2002) and linguists (e.g. Fauconnier, 1985; Johnson, 1987; Lakoff, 1987; Lakoff & Johnson, 1980; Langacker, 1987, 1990; Talmy, 1983). Despite the many differences, all authors agree that our conceptual capacities utilize perceptual resources: concepts are couched in representational codes that are specific to our perceptual system (the modal-specificity hypothesis). The *embodied* (or *grounded)* view assumes that concepts arise directly from representative brain states. Concepts (brain states) are then reactivated, i.e. they are *simulations* of the perceptual and interoceptive experience. The simulation is then available for language and thought processes.[34] The key difference between the two perspectives is that the disembodied view of concepts assumes that concepts are mental representations fundamentally unlike what they represent. Thus, critically, perceptual and interoceptive brain states are not constitutive of concepts. For embodied cognition proponents, simulations, in contrast, are analog presentations, in the sense of re-presentations of perceptual and interoceptive experiences; they are directly grounded in the body-based and subjective perceptual states that give rise to them.

[34] Most versions of grounded cognition (e.g. Barsalou, 1999; A. Clark, 1997) assume that although brain states are constitutive of concepts, concepts are nevertheless representations, and thus distinct from brain states (e.g., sensorimotor activations). Barsalou (1999) refers to such concepts as "perceptual symbols": on this view, perceptual symbols are stored traces of prior brain states that can be reactivated, or simulated, for purposes of language and thought.

Our discussion on the nature, format, and structure of concepts could go on for many more pages. Let us briefly summarize what kind of general view about concepts seems to underpin many approaches in psychology and in the cognitive sciences. Smith and Medin (1981) opened *Categories and Concepts* by declaring that concepts give human experience stability and that mental life would be chaotic without concepts. Murphy (2002) expresses a similar feeling arguing that concepts hold our mental world together: concepts are mental representations of classes of objects in the world and that they tell their holders what things there are and what properties they have. Carey (2009) indicates that concepts are units of thoughts that are the constituents of beliefs and theories. Solomon, Medin, and Lynch (1999) offer that concepts are building blocks of thought and that they, among their functions, support classification, inference, and conceptual combination. Taken together, such remarks suggest that concepts are stable units of knowledge in long-term memory that represent meaningful sets of entities in the world and provide the elements out of which more complex thoughts are constructed (cf. Malt et al., 2015).

Finally, we can turn to our main point of interest: how do concepts (simple or complex thoughts, described as above) relate to language? We have made it so far without even mentioning language explicitly. However, as we will see briefly, there is a very strong connection between theories of concepts and theories of meaning in the psychological literature we reviewed. Once more, we will restrict our discussion to some very general considerations, leaving the analysis of the specificities of each proposal to subsequent chapters.

A first issue that has been variously addressed in the philosophical and psychological literature is if we can have thought without language in the first place (for a review cf. Margolis & Laurence, 2019). We will not engage with this question as psychologists have a pretty straightforward answer: yes, of course we can think without language. Surely, we know that non-human animals engage in behavior that is sophisticated enough to warrant ascription of concepts. For instance, human infants (prior to language acquisition), chimps, and even parrots can group similar things together, i.e. they can classify objects. In our everyday life, we have the strong feeling that there are many concepts that we are not able to express in language, and so on.[35]

[35] We have already discussed the issue of linguistic determinism in the firtst introductory chapter. Of course, this issue is much more profound, and it has a long history in philosophy. However, we will not provide an in-depth discussion as it would take us far from our main issues and in current approaches in psychology and in the cognitive sciences it is pretty much established that there can be thought without language.

A different question is whether we need to integrate a theory of concepts with a theory of language or what the relationship between concepts and language might be. Saying that one can possess concepts without language does not imply that language plays no role in our conceptual ability. Interestingly, many psychologists, often arguing against more or less well understood version of the relativity hypothesis, state (almost in passing) that "the only respect in which language clearly and obviously does influence thought turns out to be rather banal: it provides us with most of our concepts" (Devitt & Sterelny, 1987, p. 178), a view far stronger than that of even the most pro-Whorf researcher. Similarly, Prinz (almost in passing again) states that "language is, first of all, a dominant means of learning new concepts [...]. Moreover, concepts that get lexicalized are often more salient and easier to learn. Language also aids in using concepts" (2002, p. 21). Nonetheless, even if language may be implicated (so some extent) in building our concepts, Prinz suggests that a theory of what concepts are can be achieved without mentioning language. The often-tacit (made explicit by Prinz) assumption in the study of concepts is that concepts are language-independent, i.e. they are constructed prior to and independently from language.

A variety of approaches in psychology and in the cognitive sciences share the view that cognition (in this case thinking and the conceptual structure thereafter) arise independently from language; they are either (mostly) innate and the words of a language simply labels them (J. A. Fodor, 1975), or they are acquired through (for example) the senses and the words of a language (not as simply as in Fodor's view) map onto them. Words reflect the content and the structure of our conceptual organization. According to a fairly big share of approaches in psychology and in the cognitive sciences, meaning is conceptual. A familiar and widely made assumption in non-truth-functional semantics is that meanings are, or are subparts of, the conceptual representations deployed in general cognitive processes. A very broadly accepted definition of meaning is: "the aspect of words that gives them significance and relates them to the world", and "words gain their significance by being connected to concepts" (citations from G. L. Murphy, 2002, p. 385; cf. Jackendoff, 1996; Wierzbicka, 1996; Chomsky, 2000; V. Evans & Green, 2006). Words map onto concepts (or sub-portions of them): "humans invent words that label their concepts" and "these linguistic categories and structures are more-or-less straightforward mappings from a preexisting conceptual space, programmed into our biological nature" (Li & Gleitman, 2002, p. 266) (Li and Gleitman, 2002, p. 266) (see figure 6 for a representation of the conceptual view of meaning).

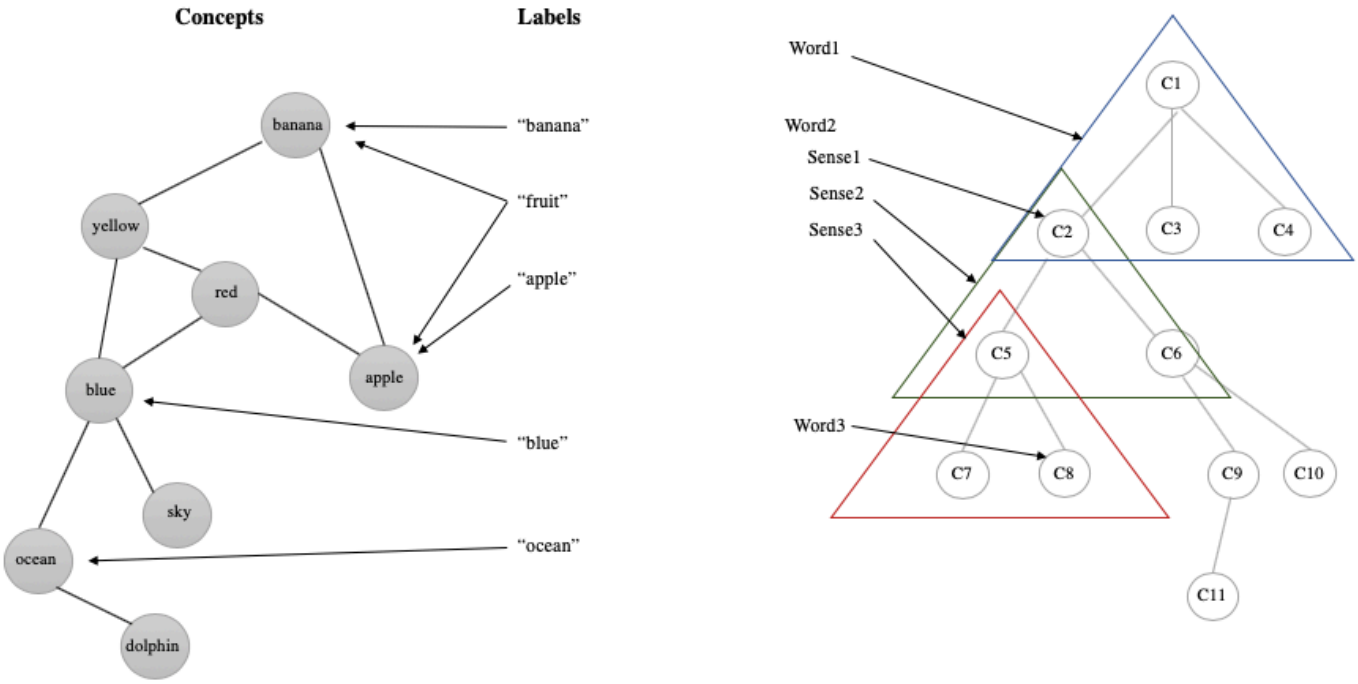


Figure 6. The word-as-mapping view of meaning. On the left: a representation of how words label pre-determined concepts which already provide the content of our knowledge. On the right: the meaning of words consists in subparts of lexical knowledge that are picked out by the lexical item. Words can be composed by word senses that map onto concepts or set/sub-portions of concepts and contribute in the mapping structure of a given lexeme.

Different theories take then different stands of the format, the source, the nature, and so on, of the underlying notion of concepts, on which word meaning is then connected (we will come back to the specificity of these theories when discussing the structure of our conceptual/lexical knowledge in chapter 3). We are at the heart of cognitive linguistics (and more specifically, of cognitive semantics). Murphy's chapter on word meanings in *The Big Book of Concepts* can be taken as a clear, textbook-like explanation and summary of the leading view in understanding the relationship between thought and language. For example, Murphy shows that general properties of concepts can be found also in the use of words and argues that this is evidence that the words are represented via those concepts: "if words are not represented in terms of concepts, then they must be represented in terms of something else that just happens to have the exact same properties as concepts. By Occam's razor, I will conclude that word meanings are represented in terms of concepts" (2002, p. 394).

Conceptualist identifications of meaning have recently increasingly come into question both within and outside cognitive linguistics. Within linguistics, the explanatory adequacy of a

concept-based semantics is undermined by a spectacular lack of agreement among investigators about what the concepts underlying meaning are like (for a review and literature selection cf. Riemer, 2013). Similarly, outside cognitive linguistics, the first problem that pops up in defining word meaning in terms of concepts is that we do not have a commonly accepted definition of concepts to begin with. This lack of precision in the notion of concept has a direct consequence for attempts in conceptualist linguistics to analyze the semantics of natural language. Furthermore, the literature on word meaning has not been directed towards distinguishing theories of concept representation (and vice versa). As a result, the phenomena that have been discussing are generally not very probative in evaluating different accounts of how concepts are represented. For example, as Murphy puts it, "although typicality effects have been found in linguistic data, in sentence comprehension, speech production, and rating studies, these studies have not investigated whether people representations of word meanings are exemplar-based or some kind of prototype representation [...]. Often such studies assume a prototype representation, but it seems likely that exemplar models could explain most of the data just as easily" (2002, p. 440).

Another line of criticism has pointed out that the claim that conceptual representations constituting linguistic meaning are the same representations that figure elsewhere in cognition is unjustified if we cannot provide an account of concepts *independent* of our account of language. A remark is here is order. Although we have undertaken the debate of the relationships between language and thought from a different perspective compared to the previous chapter, we have nonetheless reached a similar conclusion: the main (at least, methodological) issue in the study of the relationship between language and thought is the very possibility of separating the two. We saw how this criticism was powerful against researches in linguistic relativity: we cannot say that language influences thought if we cannot provide non-linguistic (hence, purely conceptual) proofs of such influence. The critique is similarly powerful when addressed to theories that view language as a mere mapping device:

> without an independent idea of what concepts outside language are like, we lack the means to check the claim that linguistic meanings are concepts. In order to escape this criticism, a naturalistic semantic theory must look outside language for an understanding of the nature of the concepts with which it identifies meanings, and then ask whether the explanation of language requires the postulation of the same or of different kinds of concepts, or the postulation of concepts at all (Riemer, 2013, p. 3).

For example, decompositional accounts of word meanings in terms of primitive concepts seem not to avoid the problem of circularity mentioned by Riemer. Since a definition can be proposed for practically any word, one can simply interpret the terms of the definition as concepts, and see in the structure of definitions clues to the structure of the word's conceptual representation. The conceptual system as it emerges in this kind of auto-exemplificational analysis simply mirrors the object-language meanings it is supposed to explain. The conclusion that *keep* means KEEP (J. A. Fodor, 1998, p. 55) in itself tells us nothing about the psychological architecture supporting language.

However, despite these and other criticisms that cognitive accounts of meaning necessarily have to face, the explanatory power of conceptual explanations is outstanding; there is, surely, something broadly right about that claim that semantic aspects of language are connected in some way to cognitive agents' knowledge structures. As we have hinted before, the relation between language and thought should be understood (or so we will argue) as a bidirectional relationship, also at the level of the explanatory power that a theory of concepts has for a theory of language, as well as the opposite. In the cognitive literature, there is now a solid body of research that argues both theoretically - as well as by providing empirical evidence - for the fact the language can be viewed as a window on thought, i.e. that by studying language we can reveal aspects of our cognitive and (more specifically) conceptual structure. The language-as-a-window-on-though perspective is grounded on the idea that (at least a good portion of) the meaning of words is given by the connections of words with concepts. In the next section, we will review some of the arguments that point in that direction and we will try to show how, although being to some extent compelling, they cannot exhaust the study of meaning.

2.1.3 Beyond language as a window on thought

To summarize, in many cognitivist accounts of language, concepts are independent and prior to language; words label (sub-portions) of concepts or of our conceptual structure more broadly; the function of language is that of transferring thoughts from speakers to hearers. On the basis of this view, what does the study of language allow us to reveal about the nature and structure of our concept?

One first aspect of thought that such a view of language allows us to reveal is the conceptual property of *invariance*,[36] i.e. the idea that concepts are, in essential respects, identical from one individual to the next (e.g. J. A. Fodor, 1998). For example, Prinz (2002) argues (very briefly and without developing the argument) that the fact that we can convey meaning by talking is an index of the fact that our concepts are "public" (hence, invariant). The general outline of the argument is this. According to the standard picture, people understand each other's words in virtue of the fact that they associate the same (or nearly the same) concepts with those words. If no two people associate the same concepts with their words, then communication is impossible. Therefore, concepts must be sharable. Clearly, this issue is a matter of intense debate and the shared, public nature of concepts is not straightforward at all (cf. Laurence & Margolis, 1999; Riemer, 2013). However, we will leave it like this since we want to draw attention to something different and more relevant to our discussion.

Language seems to reveal the organization of our conceptual structure. As we have seen, the prevailing assumptions seems to be that many concepts can be identified because they are revealed by words. For example, Smith and Medin (1981) used English nouns such *as hat, fish, triangle, table,* and *robin* to identify concepts. And similarly did many other researches (e.g. Lakoff, 1987; Rosch & Mervis, 1975). As Murphy plainly points out, "I can talk about children learning the concept sheep, say, but I can also talk about their learning the word *sheep*. I would take as evidence for a child having this concept the fact that he or she uses the word correctly. On a purely intuitive basis, then, there appears to be considerable similarity between word meanings and concepts. And, in fact, much of the literature uses the terms interchangeably" (2002, p. 385). After all, if entities in the world fall into natural groupings according to their shared properties, then it also makes sense to take English words to capture these groupings and to be associated with a coherent mental representation. A consequence of this view, as Malt et al. (2015) argue, is that English can be taken as the language of research without concerns, because other languages should have words that capture the same groupings and are, therefore, associated with the same concepts.

[36] The property of the "invariance" is actually a requirement of any theory of thought inasmuch as we want to assure communication. A traditional objection to the conceptualist identification of meaning is that it makes meanings private mental entities which cannot be shared. Since communication is thought of as the transfer of the same thoughts or concepts from speaker to hearer, the hypothesis that meanings are essentially private might seem to require us to jettison our view of communication. However, as Riemer argues, "as long as we assume that there is some overlap in the mental content expressed by words, we can retain the traditional view of communication while upholding a conceptualist theory of meaning" (Riemer, 2015, p. 37).

Forasmuch as there would be no reason to suppose that the noun *chair* in a different language (say, *sedia)* would capture a set of objects sharing properties (such as having a seat, a back, legs, and supporting a human in a particular position) different from those of English.

But there are different ways, definitely more compelling ways, of framing the idea that language is a window on thought. Important insights come from the embodied literature. As we have previously mentioned, the general claim of embodied/grounded theories is that cognition (and concepts) are grounded in our bodily and sensorimotor states, which subserve language. Embodied/grounded theories of concepts assume that linguistic meaning is equivalent to conceptual representations. Hence, if language is subserved by concepts grounded in multimodal brain states, we should find evidence of grounding effect is language. Consequently, by examining language we gain a direct vista onto the nature and mechanisms of the human conceptual system, upon which language depends (Talmy, 2000 was one of the first who saw that language encodes embodied concepts; in contemporary research, this argument is strongly made by V. Evans, 2009, 2015a, 2015b - but we will deal with Evan's perspective in more details later, since it provides a more complex and multifaceted account of language). Perhaps the clearest evidence of how language reflects the structuring and structure of our thought has been highlighted in the work of Lakoff and Johnson (1980, 1999) on conceptual metaphors. Conceptual metaphors appear to work by recruiting structure from sensorimotor experiences in concrete domains in order to structure representations in other (often more abstract) domains; the domain of time is systematically structured by establishing long-term stable links to the domain of motion through space. This is clearly reflected in the way we talk about time, providing evidence for different conceptual metaphors: "Christmas is approaching", "The time for action has arrived", "The summer just zoomed by" (examples from V. Evans, 2015a, p. 73).

Similarly, if language provides a window on thought, and thought is grounded in sensory-motor experience, then we would expect the structure of language to reflect and reveal such grounded components. Evans builds the following argument:

> the foundational domains of experience are those of time and space. [...] the stuff that makes up the domain of SPACE is matter, of which two broad types can be distinguished: discrete entities (e.g. objects) and mass entities (e.g. fluids). The stuff that makes up TIME is that of action. As with matter action can also be broadly subdivided into action that is discrete, or bounded, and that which is unbounded. These distinctions are captured by a fundamental design feature of the grammatical system of

> language. The lexical classes noun and verbs appear to be common to all, or at least nearly all, of the world's languages. Objects from the domain of SPACE appear to be the prototype for the lexical category noun, while the notion of a dynamic and ongoing process in the prototype that underpins the grammatical category verb (V. Evans, 2015a, p. 105).

His conclusion is that grammar embeds embodied experiences in its very make-up. Experiences relating to time, space, attention, force dynamics, and perspective point are clearly reflected in language and give it its scaffolding.

Although this perspective may very well account for a broad range of data and intuitions about the source of meaning in language, it is nonetheless clear that some fundamental aspects of what meaning in language is and of what language does seem difficult to encompass within the idea that language merely reflects thought. The fact that language may play a more or less prominent role in "programming the mind" (Lupyan & Bergen, 2016) is a conclusion that has been reached by a number of leading cognitive scientists. For example, Barsalou (2005; Barsalou, Santos, Simmons, & Wilson, 2008; see also V. Evans, 2015b for review) has suggested that the function of language is to provide an executive control function, operating over grounded concepts in the conceptual system. The idea is that language provides the framework that facilitates the composition of concepts for purposes of communication. Hence, language adds something to our cognitive abilities: it provides us with a means of controlling our mental representations by manipulating embody-sensorimotor representations. What aspects of language provide us with this added ability? Mostly grammar, or so it has been argued (e.g. Bergen, 2012; V. Evans, 2015a, for a review cf. 2015b). The manipulation of sensorimotor representations is achieved by language consisting of a grammatical system, with words and constructions cuing activations of specific body-based states in the brain. As Bergen (2012) summarizes it, grammar combines and thereby constrains the contributions that individual words make to embodied simulation; grammatical constructions covey meaning that is directly fleshed out in concrete components of an embodied simulation; grammar modulates embodied simulations, indicating not what to simulate, but how to simulate it.

One fundamental aspect on which researches have largely disagreed is whether language conveys meaning of its own. On the one side, researches have argued that language has no semantic properties of its own, independent of the simulations produced by grounded concepts in the conceptual system (Barsalou et al., 2008; A. Glenberg & Robertson, 1999). The role of language

is to provide structure (and not content), which aids the assembly of perceptual states in the construction of meaning. However, it has been pointed out that the grammatical subsystem appears to encode semantic content independently form the conceptual system (Bergen, 2012; V. Evans, 2009; V. Evans & Green, 2006). For example, representations associated with grammatical structure, so called closed-class elements, appear not to relate, in a straightforward way, with perceptual representations, and yet they are meaningful.

An outstanding attempt to delineate a language-specific level of meaning is that of Evans's (2009, 2010, 2013, 2015a) proposal of distinguishing between two sets of qualitatively different concepts (and structures): one in the conceptual system, older in evolutionary terms and grounded in bodily-sensorimotor experiences, and one in language, which evolved to provide an executive control function in harnessing concepts in the conceptual system. Evans argues that the distinction between the conceptual and linguistic systems relates to concepts that are analog in nature (they are directly grounded in the experiences that give rise to them), on the one hand, and those that are parametric in nature, on the other.[37] Parametric (schematic) concepts (i.e. linguistic concepts) facilitate access to analog concepts in the process of meaning construction. Of what kind are these language specific representations? There appears to be a type of semantic representation that, arguably, is unique to the linguistic system. As we have seen when mentioning Berger's perspective, grammar provides information relating to how a simulation should be constructed (in the service of linguistically mediated meaning construction for the means of communication). After all, the grammatical organization of the sentence entails that the first entity is the agent and the second entity the patient: the first entity is performing an action that affects the second entity. This level of semantic representation derives exclusively from language, rather than from conceptual structure, providing an instruction as to the relative significance, and the relation that holds, between these two entities. In short, closed-class elements, and the grammatical configurations in which they reside (which are themselves closed-class elements) involve semantic

[37] The distinction of linguistic concepts and conceptual concepts is preliminary to the development of a theory of how the two contribute in different ways (and with different functions) in the process of meaning making, a theory of lexical representation and semantic composition referred to as the theory of lexical concepts and cognitive models (LCCM). The LCCM theory framework provides a basis for understanding the respective contribution of each distinct concept type in facilitating linguistically mediated meaning construction, which is complementary, albeit orthogonal to, an account of knowledge representation (concepts). Here, we will not engage in the discussion of the LCCM theory directly as the distinction between conceptual structure and language structure is more relevant for the purposes of our present argument (for LCCM theory presentation and review cf. for example, Evans 2015).

content, albeit of a highly schematic sort (V. Evans, 2009; A. E. Goldberg, 1995, 2006). Grammatical meaning is schematic in nature and it provides structural information:

> While analogue concepts – directly associated with open-class words, and housed in the non-linguistic conceptual system – convey the *what* of a simulation, the closed-class elements encoded by human grammar – by language – provide the packaging that allows us to nuance how the analogue concepts are presented. (V. Evans, 2015a, p. 235)

Furthermore, Evans argues that also open-class words convey schematic-parametric (hence, linguistic) meaning. Parameters are abstracted from embodied (perceptual and interoceptive) states, filtering out all points of difference to leave highly image-schematic content. The example he provides is based on another observation starting from grammar; words can be divided into different "lexical classes": nouns, verbs, adjective, and so on. Nouns are taken to refer to things, while verbs concern relations that evolve through time. Adjectives seem to denote properties of things. Different lexical classes package analogue content (multimodal information found in the conceptual system) in different ways. In the example drawn from how these two sentences convey the same analogue representation (that part of the conceptual space corresponding to the color spectrum usually identified with "red") - "Treat redness with Clinique urgent relief cream" and "Treat red skin with Clinique urgent relief cream" – the content "red" is package into different ways, giving rise to different simulations. In the case of "redness", the word encodes the parament "thing" (color is reified as an entity); in the case of "red", the word encodes the parameter "property", i.e. whatever is being pointed to in the conceptual systems, it has to be identified as a property of something. Hence, highly schematic notions such as "property" and "things" are the parametric concepts that are proper of the language system and based on which we can package richer, analogue meanings in the conceptual system (for a discussion see Evans, 2015, Chapters 8 and 9). This type of representations is unique to the language system.

Let us pause for a second and make the point of our discussion so far. On the one side, language can indeed be viewed as a window on thought: it provides us a mean by which we can pick at the conceptual structure and conceptual processes that underlie language. But the extent to which language can be seen as a window on thought needs to be reconsidered. We have briefly presented Evans's view on this: while he maintains that language shows us that meaning is embodied, and that language reflect the structure of our conceptual system, language also exhibits a level of language-unique meaning, that of parametric concepts, which allow us to package our

conceptual knowledge in different ways (if anything, for being communicated). Conceptual concepts and linguistic concepts are intertwined in the processes of meaning-making. Language provides us with a means of harnessing the conceptual concepts in the process of meaning-making by elaborating, nuancing, and combining them in complex ways.

In the course of our work, we will build our argument towards the possibility of identifying another level of language-specific meaning. We will try to show how language not only allows us to package non-linguistic meanings in language-specific ways, but how it also provides us with part of our conceptual knowledge. Our argument starts from the intuition that language provides a language-specific organization of meaning beyond the organization provided by the grammar of a language. Certainly, one source of our knowledge and our shared concepts is the common evolutionary history of humans that gave rise to similar bodies and similar sensory-motor capacities. However, our shared biology and environments drastically under-determine the possible concepts human may end up having (for similar arguments cf. Lupyan & Bergen, 2016; Malt et al., 2015). In the next paragraphs we will draw from positions in linguistics and we will try to show how the words of language cannot simply map onto non-linguistic concepts, inasmuch as they carve up reality in ways that are not always consistent with the joints of nature or with (supposedly) pre-linguistic concepts. As we have argued in the previous chapter, this becomes rather obvious if we take into consideration language diversities in the study of meaning. In fact, language diversities call into question the very idea that language may simply map onto non-linguistic concepts. Before turning to the main act of this chapter, i.e. to the discussion of if and how language carves nature at its joints, let us outline some preliminary reflections of the study of meaning in general and, more specifically, on the notions of *semantics* and *meaning* in the first place.

2.2. Interlude: approaching a definition of meaning

In approaching the kind of enterprise described above, i.e. in discussing language and linguistic meaning as the point of intersection between language as a product and language as an aspect of cognition, we are stepping onto a slippery terrain: we are entering that labyrinth of notions, theories, and discussions that address (more or less explicitly) the problem of semantics (as a field of study and as the subject of a field of study) and of meaning (meaning as a general term, meaning as linguistic meaning, etc.). Our aim is not, and cannot be given the stage of our research, that of an exhaustive reconstruction of the history of semantic theories. In this section, we aim at clarifying our theoretical/foundational grounds in order to delimit the field of inquiry we are about to explore and on which we will try to build upon. Nonetheless, even after this restriction, the task does not feel any easier than before. Meaning seems hard to grasp and a unified definition of the notion of meaning has not been established yet. According to the definition of meaning one adopts for his or her research, the set of phenomena the fall into the study of semantics varies dramatically. Hence, we will spend quite a bit of ink trying to outline the fundamental assumptions underlying our work and their groundings in the history of linguistic ideas, situating our research in a broader theoretical perspective devoted to the study and understanding of the nature of human signification.

2.2.1 Semantics: introductory notes

Introduced in 1883 by Michael Breal, the term *semantics* went from being a technical term limited to the study of linguistic meaning to a common word both in its various applications in disciplines other than linguistics as well as in everyday language and standard dictionaries of many languages.

The case of re-interpretation and operationalization of the term *semantics* in various disciplines is particularly significant for the purposes of our discussions, since it shows how the study of meaning is, even nowadays, far from being a unified discipline. Let us just mention some interesting examples. One of the first approaches in distributional semantics carries the world *semantics* directly in its name: *Latent Semantic Analysis* (LSA) "is a theory and method for

extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text" (Landauer, Foltz, & Laham, 1998, p. 259). Already in the first sentences of the paper we can intuitively see an underlying assumption about the characteristic of the study of semantics and how meaning is implicitly defined. In the discipline of Artificial Intelligence, the sub-field of Natural Language Processing is exponentially growing on a daily basis and the implicit notion(s) of semantics underlying the many approaches are often diverse and not of straightforward reconstruction (Manning & Schutze, 1999). Probably even more complex is the use of the term semantics in psycholinguistic and psychological approaches, in which the notion of semantics (and more often of semantic knowledge or of semantic representations) has shifted to indicate our general, linguistic and non-linguistic, knowledge about the world (see for example: Lupyan & Lewis, 2017; Vigliocco et al., 2009). In mentioning these studies and in pointing out that they implicitly rest on more or less specific notions of semantics, we by no means intend to underestimate or diminish neither their theoretical work nor their methods. Our intention is indeed the opposite; given the importance and the ground-breaking role of such research paradigms (to which we will refer again and largely discuss later) we are endorsing the idea of the need for an integrated theoretical and terminological framework for the study of meaning.

Such an enterprise has been the focus of much of the theoretical work of the Italian philosopher and linguist Tullio De Mauro, who extensively worked towards a "unified" and "integrated" notion of semantics as a "general theory of signs" (De Mauro, 1982b, p. 473) that could encompass different views on meaning and methodological approaches within a strong formal theoretical structure (De Mauro, 1982b, 1982a; Gensini, 2019). De Mauro argues for a requalification of semantics within the broader field of semiotic studies, i.e. he points out and to the fact that linguistic meaning ought to be studied by means of a comparative approach between human language and other systems of communication. Such a contrastive perspective allows to grasp the uniqueness of human language in relation to other systems like animal languages, artificial languages (like calculus, signages, and so on) and other natural-historical languages (as signed languages), by means of analyzing the different ways in which these (and other) systems organize the realm of contents (Gensini, 2019, p. 8). Thus, De Mauro proposes a distinction within the study of semantics between *semiotic semantics*, i.e. the study of how signs organize the realm of contents, and *linguistic semantics*, i.e. a sub-field of the semiotic semantics that is interested in the study of the whole of the phenomenon of signification of historical-natural languages in its

distinctive traits while remaining anchored to the principles and understandings of the broader field of semiotics. When compared to other systems of communication, historical-natural languages reveal their peculiarities: the double articulation of the linguistic sign, the creativity of language, the non-total predictability of the extension of the senses made possible by their combination, the self-reflecting ability of the humans' linguistic activity are some of the fundamental and constitutional peculiarities that make languages what they are (De Mauro, 1982b; Gensini, 2019, p. 9). Our contribution situates itself in this theoretical framework and it is especially grounded in the grand scheme of those peculiarities that make human languages a powerful cognitive and communicate tool and that seem to find their maximal actualization in the study of linguistic meaning.

2.2.2 Meaning: bits and pieces of linguistic theories

The AI, psycholinguistic, and linguistic studies briefly mentioned above seem to fall under the category of *linguistic semantics*, which studies the phenomenon of meaning in natural-historical languages. But this consideration doesn't tell us much by itself. Even within the theoretical framework of a *linguistic* semantics, much of what ends up being in the lens of the researcher depends on the definition of meaning we assume.

The notion of meaning is central in understanding the nature of language and how it works; but despite the recognition of its centrality for a theory of language, meaning reveals itself as a complex entity. This complexity is evident in the lack of a shared and commonly-agreed-upon definition not just within different disciplines directly involved with the study of meaning, such as linguistics and philosophy, but also among researches coming from areas not directly concerned with the study of language. The first observation in attesting this complexity is that meaning involves multiple components that are not reducible one to the other; to this regard, the famous semiotic triangle proposed by C. K. Ogden and I.A. Richards in *The Meaning of Meaning* (1923) comes handy; it distinctly shows that meaning is the place of the interaction between language (*expression)*, thought (*content*) and reality (*referent*). Following various critical reconstructions of the history and developments of the notion of meaning (Gambarara, 1999; Violi, 1997), we can understand the numerous definitions of meaning as means of putting different weight on different sides of the triangle; 1) the meaning of an expression is the "entity", the thing, the state of affairs

to which a certain expression refers to (referential semantics); 2) the meaning is the concept, the mental image, to which the expression in linked in our minds (cognitive semantics); 3) meaning is the relationship that each linguistic unit, understood as the union of *signified* and *signifier* (content and expression), entails with other linguistic units (structural semantics).

Although this cannot be the place of an exhaustive discussion of the strengths and weaknesses of the different definitions, we will sketch some general observations useful to our discussion.

The referential and the cognitive views of meaning seem to share a common underlying assumption; i.e., language is fundamentally a labelling device that maps onto a pre-organized external or mental reality. On the one side, according to the referential approaches[38], linguistic meaning is anchored onto the external reality, understood as sum of extralinguistic "state of affairs". Thus, referential accounts of meaning underestimate or completely overlook the autonomy of the specifically systematic dimensions of meaning. On the other side, as we have extensively argued in the previous sections, cognitive perspectives on meaning seem to emphasis a sort of "cognitive priority" of the organization of thought, which turns into the idea that words simply map onto concepts by means of what we could call a labelling-function. Some of the problems that arise from these accounts (Basile, 2001; Lupyan & Lewis, 2017; Saeed, 1997, pp. 30–31) point to the fact that they hardy actually take into consideration the problem of linguistic meaning or restrict the phenomena they observe to specific sub-categories of the lexicon and ignore the whole of the ordinary language.

On the contrary, what we are interested in, and probably what a *linguistic semantics* as opposed to a *formal semantics* (De Mauro, 1982b) should do, is to delineate a theory of linguistic meaning based on the variety of everyday uses of language and that would allow us to understand not just how language represent things or facts but also how language, and linguistic meaning specifically, allows us to deal with the continuous fluctuation of the communicative practice as well as with the variety of the senses and of the individual and idiosyncratic representations of the meaning of words (De Mauro, 1965).

[38] For the sake of the argument and of the line of inquiry we are developing, we are here condensing in their most general terms a broad range of different approaches, theories, and discussions that often do not show the uniformity we are implying in our statement. For a review of the philosophical-analytical approaches to meaning see among others Violi (1997) and for a discussion of the relationship between meaning and referent in the analytic philosophy and in the cognitive sciences see Mazzone (1999).

We will now turn to the third option outlined above: structural semantics[39]. According to the structuralist perspective, the meaning of a unit of language can be described based on its relationships with all other linguistic units in the system. Consequently, linguistic meaning is displayed in a web of relationships and dependences internal to language and independent from each and every determining factor external to the system. This general claim is often taken to be the foundational ground of all research in structural semantics, from its foundation in the work of the linguist Ferdinand De Saussure, to Hjelmslev's notion of system, to its development in the subsequent approaches in structural semantics (Violi, 1997, p. 30). According to the structuralist reading of the Saussurian themes, language is an abstract, autonomous, self-sufficient entity that operates as a categorization mechanism by giving form to the amorphous mass of thought; linguistic meaning, and consequently the study of it, is independent from factors external to the system, i.e. from the perceived world and from non-linguistic external as well as conceptual reality. The view of meaning put forward by the structuralist approach seems to be fundamentally non-referential in the sense that language does not simply map onto a pre-organized external (or internal) non-linguistic reality.

This anti-referential (and anti-psychological) view of meaning stems from the very definition of the linguistic sign proposed by Saussure. In the Saussurian perspective, the linguistic sign is an entity made of two faces (the *signifier* – the expression - and the *signified* – the content of the sign) that shape each other reciprocally in the process of signification by carving up the

[39] Structural linguistics is not a unified paradigm of research, rather a multiplicity of approaches that declared to have been inspired by the Saussure's *Course in general linguistics* and that mostly focused on some Saussurian dichotomies such as *langue-parole, synchronic-diachronic approach, syntagmatic-paradigmatic*. The history of the *Course in general linguistics*, both in its writing and its reception, is not easy and straightforward. Given that this cannot be the place for an exhaustive summary of it, we will here limit ourselves to indicating some renowned sources for historical and critical readings (De Mauro, 1967a; De Palo, 2016; Lepschy, 1997, 1998). Nonetheless, it is necessary to briefly mention a recent development in the history of the structuralist approach folllowing the publication of the unpublished essays of Saussure (Saussure, 2006). Two of these essays (published under *De l'essence double du langage*) have been fundamental in presenting to the public a new side of the Saussurian's theories (De Palo, 2016, p. 25). In the course of our discussion we will mostly discuss and endorse the interpretation of Saussure's work put forward by the "Roman" school of Antonino Pagliaro and Tullio De Mauro. In this renewed reading of Saussure some theme stand out as prominent: the problem of the linguistic identity, the notion of meaning as an open class of senses and significations, the attention to the speaker and his/her knowledge and representations of the world, the limits of the arbitrariness of language imposed by, for example, our cognitive limits (De Palo, 2016, p. 26). We will use the terms *structuralism* and *structuralist* to refer to those approaches that have been inspired by Saussure and, conversely, we will refer directly to Saussure's work (mostly as read by De Mauro) when considering the renewed interpretation of the Saussurian thought. This distinction is fundamental since the renewed reading of Saussure's work shows different interpretative paths and, to some extent, is not subjected to the same objection to which the structuralist approaches were subjected.

amorphous substrate of the mass of concepts and of possible sounds. The linguistic sign states its independency from the non-linguistic world (the referent) and from the subjective, physical and psychological world of the user (i.e. from the mind of the speaker); the linguistic and semantic units of language have been considered as purely formal (differential) and correlative entities deprived from their subjective basis (De Palo, 2016, p. 35). This linguistic articulation of the space of concepts and of sounds is radically arbitrary, i.e. it is not motivated by any natural or logical reason. Therefore, the distinctions made by and within the language system seem to have little or nothing to do with the intrinsic characteristics of the space of the concepts. It is the process of signification internal to the system that shapes the concepts, which are not pre-defined. According to the structuralist reading of Saussure, the meaning of the linguistic unit (the *value*) determines itself in a web of oppositive relationships internal to each historical-natural language and, therefore, it is independent from the way in which humans might perceive or conceptualize the external reality in a non or pre-linguistic way. Hence, the structuralist approach to meaning, from its origins to it later developments, seem to lead towards an autonomous view of semantics[40] and to the possibility of studying meaning as separated from the referential values of words and from the biological constitution of human beings[41] (cf. Casadei, 1999; Violi, 1997).

The anti-psychological and anti-experiential nature of the Saussurian theory has been and still is a matter of debate. The themes around which we can identify the anti-psychological nature of Saussure's thought can be summarized as follow:

> 1. il segno, e di conseguenza il significato, non sono di natura introspettiva, ma sociale; 2. il significato non è un concetto, ma è puramente differenziale poiché´ è determinato dal con- testo sintagmatico e associativo; 3. la lingua non è una nomenclatura di oggetti precedentemente dati, ma un sistema di valori relazionali, una forma (un insieme di articolazioni) che si proietta sulla sostanza dei concetti (e delle idee) e dei suoni; 4. l'oggetto di studio della linguistica è autonomo, nel senso che lo studio della langue (come sistema di relazioni intralinguistiche) è indipendente dallo studio della mente e dei processi psichici (oggetto della psicologia) e dallo studio della materia fisica e fisiologica dei suoni (oggetto, per esempio, della fonetica) (De Palo, 2016, p. 36)

If, on the one side, these themes are central in the Saussurian perspective and became the focus of

[40] We shall not forget to mention the anti-semantic development of the structuralist assumptions in the work of Leonard Bloomfield (1933) and that shaped much of the subsequent North-American studies in linguistics.
[41] For a discussion of the reasons behind Saussure's need to delineate the autonomy of object of the linguistic studies see, for example, De Palo (De Palo, 2016)

the post-Saussurian structuralism, there are aspects of the Saussurian theory that allow for a much less anti-psychological view.[42]

2.2.3 Saussure's view in De Mauro's reading: systematicity revisited

We will here briefly re-propose some of the arguments introduced by Tullio De Mauro in its critical reading of Saussure's work[43] and that lead the way towards a renewed understanding of the study of linguistic meaning as well as of the notion of language as a system from the point of view of a philosophy of language focused on the centrality of the speakers and of the notion of *sense*. It is now well established that Saussure endorsed a fundamentally anti-referential and anti-denotational view of language: it is by accident that we find the linguistic sign to correspond to an object defined by the senses as *a horse*, *the fire*, *the sun* (De Mauro, 1967a, n. 129). This does not imply that the linguistic sign has no referential function at all; instead, the anti-referential view of meaning points to the fact that we cannot explain linguistic meaning by considering *only* its connections with what it refers to. In fact, on the one side, words tell us something about the world; on the other side, it is in the use of language that we connect together experiences that are perceptually diverse and by means of which we constitute the "object" that is the basis of meaning (De Mauro, 1967a, n. 129, 1967b). Therefore, the human subject that categorizes experience by means of language, and that is behind the uses of language, becomes a focal point. On the one side, we are confronted with the creativity and the uniqueness of the individual uses of words; on the other side, we ought to recognize the reciprocal inter-individual understandability of language that sheds light onto the limitations of the variability of linguistic experience.

Hence, in order to understand the mechanisms beyond language and the nature of linguistic meaning we ought to consider the dialectical relationship between *langue* and *parole*. In *Introduzione alla semantica* (1965), De Mauro explicitly reflects on an inherently problematic aspect of the Saussurian perspective, if considered from the point of view of explaining how humans communicate and understand each other by means of language. Between *langue* and *parole* there is an inherent circularity: simplifying a little, the specific acts of meaning obtain their

[42] The many ways in which the notion of the speaker as a cognitive-social subject has been addressed in the history of structuralism is the focus of De Palo's (2016) most recent volume *Saussure e gli strutturalismi. Il soggetto parlante nel pensiero linguistico del Novecento*, to which we refer for further in-depth analysis.
[43] For a comprehensive discussion of the elaboration of Saussure's work in the "Roman" school see De Palo (2016).

value only within a language system, but it is only by starting from the specific acts that the system of language is established. Neither of the two, if taken alone, is sufficient to establish the value of a given unit. How can we get out of this circularity? The answer seems to reside in the fact that language obtains its structure only by means of being continuously reshaped by a given community of speakers. It is from the observation of the concrete acts of communication that we are led to recognize that we can identify two different phonetical instantiations as "the same word" having "the same meaning" not by means of their psychological content, but rather by assuming that they represent the same value within a certain linguistic system (De Mauro, 1967a, n. 65). The language system, as characterized by Saussure and in De Mauro's reading, is inherently dependent on its existence in time and within a given historical community of speakers, the *masse parlante*. It is this rediscovered centrality of the *langue/parole* interface that opens the door to a re-emergence of cognition as a central pillar of the Saussurian view by means of the continuous work of classification and conceptualization that speakers need to engage in in order to understand each other (Gensini, 2019). The *masse parlante* becomes a constitutive aspect of language: it is because language is used within a community of speakers that it acquires its systematicity.

The second, explicit, source of De Mauro's reflections is Wittgenstein, and especially the *second* Wittgenstein of the *Philosophische Untersuchungen*. Wittgenstein, in line with Saussure, recognizes that it is only in the *use (Gebrauch)* we make of words that meaning is created. In the famous paragraph 432, Wittgenstein states that "Every sign by itself seems dead. What gives it life?—In use it is alive. Is life breathed into it there?—Or is the use its life?" (in the 1958 English translation). Meaning is a function of the use we make of words. In paragraph 23, we can read the following formulation: "But how many kinds of sentence are there? Say assertion, question, and command?—There are countless kinds: countless different kinds of use of what we call "symbols", "words", "sentences". And this multiplicity is not something fixed, given once for all; but new types of language, new language-games, as we may say, come into existence, and others become obsolete and get forgotten. (We can get a rough picture of this from the changes in mathematics.) Here the term "language-game" is meant to bring into prominence the fact that the speaking of language is part of an activity, or of a form of life." The centrality of the use and of viewing language as an activity, something inherently dynamic, sets the stage for some further reflections. De Mauro points out that the *use* Wittgenstein is talking about is clearly the use made by single, concrete, individual language users. However, as both Saussure and Wittgenstein recognize, it is

only within the tacit agreement of a given linguistic community that the (often even idiosyncratic) uses get standardized and give rise to the abstract level of the system.

It is within this theoretical background that the distinction between *signified* and *sense* finds its clarification[44]; the *senses* are the concrete and individual uses of the *signified*. Following Saussure (and De Mauro's argument), each and every *signifier* corresponds to one *signified,* but we are constantly exposed to an *ensemble of senses*. The *senses* of a word are something different from its *signified* (De Mauro, 1967a, n. 231): they are the concrete instantiation of meaning in each and every linguistic act, in a continuous negotiation between factors external to the system and the system of language. Each *sense* finds its place within sets of senses (types of use) that share at least one semiotic attribute, namely *acceptations* (cf. for example De Mauro, 1971). Acceptations are the intermediate level between the endless variability of the senses and the abstract class of the signified. When talking about the *signified* of a word, we are referring to a notion that is part of the *langue* and that constitutes an abstract class to which all the individual senses we produce or comprehend can be referred to (cf. among others Basile, 2010b). These abstract classes are what allows language to deal with both the continuity of the reality that surrounds us and that we make experience of, as well as with the communicative practice of language, which is characterized by the variability and multiplicity of the *senses*. The discussion of how languages do this will be the focus of the next section. Here we are interested in the possibilities that this renewed interest in the multiplicity of *senses* and the focus of the *uses* of language open in the study of the semantic of languages. In fact, not only linguistic meaning is not separated from the individual sensorial experiences of the external world; it is also in a continuous balance between: on the one side, being a system of categorization and at the same time a multiplicity of different uses; on the other side, the inter-subjective and social nature and the private and individual sphere. As Saussure points out in *De l'essence double du langage* (trad. ita. 2005), the relational aspects of language as they are established in the use we make of it can be better described as a *jue de signes appelé langue* (trad. ita. 2005, p. 36). Echoing the well-known Wittgensteinan formulations, the *jue de signes* – the centrality of the *uses* as the basis of what we call *meaning* – represents a more precise notion than that of *système* or *organisme* in depicting Saussure's perspective on the nature and dynamic of

[44] For the sake of the fluidity of the arguments we are proposing, we are not discussing the Saussurian notion of "creativity" (or of "lack of non creativity" (De Mauro, 1982b)), which is fundamental in order to understand the ensuing considerations about the specificity of language among other semiotic codes and about the notion of linguistic meaning .

linguistic meaning. In the notion of *use* – in the combination between the pressure of the uses and the limits imposed by the system, we see the constant extendibility of meaning in unexpected directions

Thus, the study of linguistic meaning overcomes the classical distinction between semantics and pragmatics as it became standardized in Morris's formulation (Morris, 1938), according to which semantics is the study of the relationship between the linguistic sign and its *denotatum* and pragmatics is the study of the relationship of the sign with its users. In the same way in which linguistic meaning is bounded to the fluctuating universe of the thinkable and expressible (i.e. language cannot overlook what it refers to), at the same time it is constitutively intertwined with the psychological as well as social dimensions of its users. Languages live only in the use it is made of them by their historically and socially determined speakers. A remark seems here in order. In we take Morris's distinction within a broader semiotic perspective, the distinction between the three levels should be modulated in accordance with the type of code under scrutiny. For example, in some codes the distinction between semantics and pragmatics is fundamental for the functioning of the code itself; on the contrary, in other codes (as historical-natural languages) there is a continuous interplay between the two dimensions, inasmuch as, as we have seen, one cannot be given without the other. For instance, among the various uses that can be made of language, there are some in which semantics needs to be identified with pragmatics, and others (such as in formal and formalized uses) in which the distinction has to be kept in place.

What preliminary conclusions can we draw based on the perspective outlined so far? In a certain way, the study of meaning could be better understood in its foundational grounds within a *semiotic-biotic* perspective "in virtue of which we, as human beings, necessarily undertake operations of semiosis, that is to say we try to give form, to delimit the matter of the contents" (Basile, 2016, p. 292). Considering linguistic meaning within the process of signification, the ensemble of the mechanisms by means of which humans shape the matter of contents, places the speaker at the center of the cognitive and signifying activities, among which the process of naming gains a fundamental role (Basile, 2005, 2016). At the same time, it entails to consider the speakers as constitutionally embedded in the social-biological environment in which they live and construct their experiences. In De Mauro's reading of Saussure's perspective on the constitutive interrelation between *langue* and *parole*, we find the ground to gain a view of the language as a system without dismissing meaning as something psychological and (inherently) mental. If we ought to simplify

this point for the sake of the argument, linguistic meaning arises from the constant interrelation between the concrete uses that individual speakers make of language and the relational-systematic aspects of meaning. Hence, meaning is grounded on subjective basis; but the subject of language (the speaker) has to be thought of not just as a physiological-psychological unit, but rather also as part of a community, and in this specific case, of a community of speakers.

Did we finally come to a definition of linguistic meaning? Did we set precise boundaries for what we will call linguistic meaning in the rest of our work? No, we did not. On the one hand, as we outlined in this section, meaning is not easy (or maybe possible) to define since, as De Mauro states, semantics, as a "metalanguage", is a continuous "aporetic" condition that consists in the constant strive to capture its *indetermined*[45] and never completely calculable object (De Mauro, 1965; Gensini, 2019). On the other hand, our attempt was to sketch out the possibility that a comprehensive study of meaning should overcome strict disciplinary boundaries and set itself in the perspective of an experience-semantic-pragmatic continuum by considering how language is at the same time a physiological, psychological, linguistic, and social phenomenon.

After these preliminary introductory notes on the main notions we will be discussing throughout our work, we are now ready to finally turn to the main act of this chapter, i.e. to the discussion of how meaning may be established in language if we refute the language-as-mapping view of meaning.

[45] We will come back to the notion of "indeterminacy" as developed by De Mauro on the track of the Saussurian arguments in the next section (1.1.3.). It is relevant to point out that we are using the term "indeterminacy" and *indetermined* as translation of the Italian *indeterminatezza* and *intederminato* which indicate the property of the linguistic categories of being not fixed and logically close but rather of being characterized by open boundaries. We decided not to use the term "undetermined" (for *indetermined*) because it is most often used in the sense of under-determined or under-specified, whereas the notion of indeterminacy does not imply that linguistic categories are under-specified but instead to the fact that they are not logically closed. In general, we have tried to follow the English translations of the Saussurian fundamental terms proposed by John Joseph, an eminent interpreter of the Saussurian thought (for a recent work cf. Joseph, 2012).In our discussion, the notion of the "indeterminacy" of meaning will not receive a careful and sufficient analysis given the inherent complexity of the concept. For an in-depth analysis we refer to the volume *Ai limiti del linguaggio* (Albano Leoni, Gambarara, Gensini, Lo Piparo, & Simone, 1998) and specifically to some fundamental essays included in the volume and concerned with discussing the problematics of the notion of "indeterminacy" (Garroni, 1998) and of "vagueness" (Prampolini, 1998).

2.3. The main act: how language carves nature to its joints

2.3.1 "Raw" experiences and the "kaleidoscopic flux of impressions"

In the previous sections, we have often hinted at the idea that language does not map onto a pre-organized reality, although leaving underspecified and undiscussed alternative ways of thinking about the language-world relationship.

Let us briefly recall the assumptions underlying the linguistic relativity debate. In the Whorfian formulation, the hypothesis that the way we speak shapes the way we think was fundamentally based on the epistemological consideration that

> we dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds – and this means largely by the linguistic systems of our minds. (Whorf, 1956, p. 213)

These suggestive (and somewhat obscure) sentences gave rise to intense debates about their interpretation spanning over a window of questions that range from what actually Whorf meant, to the nature of the linguistic organization of experience, to the nature of human experience itself (Black, 1959; Kay & Kempton, 1984; Lee, 1979; Lucy, 1992).

Do we actually perceive the world as a non-organized "kaleidoscopic flux of impressions" previous to the operations of language? At a very general level of interpretation, this question addresses the long-standing issue of where knowledge comes from. At one extreme there is the idea that knowledge is built entirely from experience in the world; at the other extreme there is the conception that knowledge could be heavily innate. If knowledge comes from experience (and, at least some extent, it clearly does), then "there must be some form of filtering or focusing of attention, because not all information encountered is encoded" (Wolff & Malt, 2010, p. 4). Hence, the question is not strictly about how we perceive the world or how the world itself is organized, which mostly pertain to the realm of physiology and of natural sciences; rather, it seems to be about how we select what is salient to us in every aspect of our experience, from the realm of

perception[46], to the storage of information in memory[47], to higher and even more complicated spheres of cognitive processing. In each of these cases, language may play a role by pointing out certain ways of interpreting our experiences with the world.

In our opinion, three different questions targeting three different levels of inquiries can be identified starting from the Whorfian statement cited above: 1) is pre-linguistic experience actually "raw" experience?; 2) what kind of categorization is the linguistic categorization of reality?; 3) what is the role of linguistic categorizations in forming our experience of the world? The third question covers the entire spectrum of our research topic and hence no answer can be given upfront. The second question will be dealt with in the next section. The first need a partially separated answer that mostly lays outside of our current interests. We will only briefly review some arguments is this section, maintaining our focus on the language-thought relationship.

Is pre-linguistic experience actually "raw" experience? The answer to this question falls to some extent outside of our line of inquiry and hence will only marginally be dealt with. Nonetheless, it can be safely stated that the answer from the cognitive sciences is a clear "no". There is substantial evidence that points towards the idea that humans' categorization of experience and behavior are at least partly independent from language. Here are some examples that relate directly to our topic, but the bulk of studies from which we could draw is, of course, much wider. In the already mentioned study on the categorization of color in the Dani tribe of New Guinea, Rosch showed that Dani people, despite having only two basic color terms as opposed to eleven in English, nonetheless behave on cognitive tasks as though their color categories resembles the English system ([Rosch] Heider, 1972). The general implication of the studies conducted by Rosch was that the perception of color, and which colors are considered focal, is determined by the biology of human color perception and not by the language learned. Although these findings have been revised and counterevidence to Rosch's claims was found in further work with another New Guinea tribe (Davidoff, Davies, & Roberson, 1999), the claim that categorization may

[46] A long-standing issue in psychology of perception is the so-call "attentional blindness", i.e. in which a surprisingly salient event right in front of you may go completely unseen when unattended. This phenomenon speaks for the much broader debate about the sparsity of visual perception, i.e. the idea that we do not actually perceive all that it is in our visual scene, but that even basic perception is a matter of selectional attention (for relevant literature on the issue and references therein cf. Mack & Rock, 1998; Most, Scholl, Clifford, & Simons, 2005; Neisser & Becklen, 1975; Ward & Scholl, 2015; for the role of language in boosting otherwise unseen objects cf. Lupyan & Ward, 2013).

[47] According to many current approaches to memory organization, especially connectionist approaches, the general conceptual system is built by means of statistical associations of experiences (Burgess & Graham, 1999; McClelland, 1994)/

happened and does actually happen previously and independently from language remains overall valid. Even among studies that support the linguistic relativity hypothesis we can find evidence that shows the presence of an organization of experience that happens before the acquisition of the categories of the language. In Lucy and Gaskin study on number-marking patterns in English and Yucatec Mayan speakers (2001, 2003), the results indicate that before age 9, and long after they have acquired the Yucatec number-marking system in their language, Yucatec-speaking children display the English nonlinguistic number-marking pattern rather than Yucatec patterns shown by adult speakers of their language. Slightly different perspectives about the categorization of experience previous to language address the possibility that it is in fact thought that affects the way language itself is learned. If thought drives the acquisition of language, we might expect broad cross-linguistic similarities in the initial stage of language learning. Controversial findings characterize this domain of inquiry as well; on the side, we find evidence that children around the globe express the same semantic relations in their two-words utterances (R. Brown, 1973); on the other side, the effects of language can be seen even at this early stage, i.e. in the way children's earliest utterances reflect the word order patterns prevalent in their parents' talk (Bowerman, 1973). In her work on how deaf children, in creating their own gesture systems, deal with typological variation in how actors and patients are marked in transitive and intransitive sentences, Goldin-Meadow (2003) finds evidence for a possible universal starting point (the ergative construction) that all humans may experience before learning language. Finally, studies in early word leaning show how some aspects of conceptual experience form inevitable conflations that are conceptualized and lexicalized as unified concepts, showing a "cognitive dominance" where concepts arise from cognitive-perceptual spheres; on the other side, a "linguistic dominance" is found in those domains where the world presents conceptual bits whose clumping is not pre-ordained and languages have a say in how these bits get conflated into concepts. These distinction is reflected in the way and in the order in which children acquire meanings; children tend to acquire first concrete nouns compared to concrete verbs and, among the nouns, names of highly individuable objects and entities are acquired before those of less easily individuated objects (Gentner & Boroditsky, 2001). In conclusion, such findings, among others, seem to suggest that the way in which we categorize the world is not completely bounded to the organization of language and that pre-linguistic children show a level of categorization of experience that, in some cases, might even conflate with that of their native language.

We will not develop this point further, as it is not directly relevant to our argument here. The search for the nature and kind of relationship that is entailed between language and thought, where they differ, and where they align, does not imply that there is no pre-linguistic conceptualization/categorization nor, as we have seen, that we cannot think about side of language. On the other side, assessing that there is a level and kind of categorization of experience that is not linguistically shaped does not in itself neutralize the Whorfian claim, according to which how we carve up reality *largely* depends on our native language. We have already argued against deterministic readings of the Whorfian hypothesis (for further discussion cf. Alford, 1981; Casasanto, 2008). What interests us here is if and to what extent language simply mirrors this non-linguistic or pre-linguistic categories or if language has its (partially) own ways of carving up our experience.

2.3.2 Language diversities question the language-as-mapping view

Does language simply replicate the pre-linguistic experience we have of the world? Or, do words simply map onto a pre-organized external (or internal) reality? And by "pre-organized" we mean, conceptually pre-organized reality. In fact, we have already dismissed a purely referential view of meaning (even though both the strength and pitfalls of referential account of meaning remain a matter on enduring philosophical debate), and we have turned to a conceptual view of meaning, following both the evidence from psychology and the cognitive sciences as well as in the light of De Mauro's reading of the Saussurian perspective. Hence, the problems raised by the questions above take us back to the language-as-mapping view outlined in the previous sections. Does language simply label pre-linguistic or non-linguistic concepts?

Evidence from language diversities suggest that the answer to this question is very likely to be no. Even within the cognitive debate, it has been noted that word meanings imply categories that often do not seem to correspond to what are supposed to be cognitive useful or intuitive kinds of grouping and storing of information in the mind (cf. Wolff & Malt, 2010 for some useful examples). Furthermore, over the last decades, a substantial pool of studies has been showing that the straightforward mapping of words onto concepts is hardly defensible. Languages are not neutral in respect to the information they carry and semantic divergence between languages are not limited to cases where languages come to name different culturally determined artifacts and

institutions; languages differ largely in the inventories of words and in the content of word meanings by which they partition many domains including color, space, body parts, motion, emotion, causality (see chapters in Malt & Wolff, 2010; N. Evans & Levinson, 2009; Gentner & Goldin-Meadow, 2003a; Gumperz & Levinson, 1996b; Majid, Boster, & Bowerman, 2008; Majid et al., 2008; Majid & Levinson, 2011). These differences in words encodings are greater that they can readily be motivated by differences in experience in the physical and cultural environment.

Cross-linguistic studies now offer extensive support to the fact that the linguistic organization of experience varies across languages a lot more than previously suspected. And cross-linguistic variation occurs even in concrete domains labeled by nouns, where structure in the world might seem most likely to provide salient groupings that would be captured by the words of any language (for an example on how languages divide up ordinary household containers cf. Malt, Sloman, & Gennari, 2003; Malt, Sloman, Gennari, Shi, & Wang, 1999). Even when structure in the world does produce shared tendencies in meanings (e.g. the joints for body part terms Majid, 2010), it under-determines how any given language will map words onto elements of the world. And this diversity is not only a matter of making fewer versus more distinctions. Languages often partially or substantially cross-cut each other in the sets of entities that they group together (e.g. Bowerman, 1996a, 1996b; Malt et al., 2003).

Furthermore, besides the mentioned domain-focused studies, there has been a growing interest in recent years in evaluating cross-linguistic similarities and alignments on the basis of linguistic data analyzed by means of words' distributional properties ; Thompson, Robert and Lupyan (2018) provide a first pass at a large-scale quantification of cross-linguistic semantic alignment of approximately 1000 meanings in 55 languages finding that even ostensibly the "same" word in different languages may mean quite different things. Translational equivalents in some domains exhibit high alignment across languages while other domains show substantial cross-linguistic variability. According to this study, the domains showing least alignment pertain to human institutions, but also words pertaining to the domain of animals, common actions and physical events show intermediate levels of alignment and appear a lot more variable than in the hypothesis they just mapped onto natural categories in the world or in how our thought reflect it. Even though these observations are preliminary and need further developments, as stated by the authors, it is still the case that they point to a renewed interest in the cognitive sciences for the diversity of the semantic structure of languages and for the inadequacy of strictly referential

accounts of meaning (both if we are referring to entities in the world or to prelinguistic contents of our mental representations)[48].

Acknowledging language differences and assuming them as a key component of the study of the relation between language and thought has some evident implications in reshaping the link between conceptual representation and meaning in language (Malt, Gennari, & Imai, 2010). The various possibilities that arise from considering language differences as a key element have clear consequences for theories of mental representations in general, and for how researches would need to look for concepts.

A first hypothesis that we can now reject is the idea of a tight mapping between thoughts and words because of the underlying universality of our conceptual knowledge. Such universal concepts might come about via any of several mechanisms (or a combination of them): humans might all be driven by a set of shared needs, goals, feelings, etc., across cultures. These sources of uniformity of thought across cultures could even lead, over the course of evolutionary history, to the large stock of shared concepts being innate (J. A. Fodor, 1975). But it is by now evident that cross-culturally shared concepts leading to shared meanings can account for only a limited portion of world-to-word mappings at best. A striking finding of recent cross-linguistic research has been pervasive differences in how speakers of different languages talk about the world. The well-documented and widespread cross-linguistic diversity indicates that the situation must be more complex than implied by strong universals of human thought, coupled with a direct causal link from thought to word meaning. Above all forms of variability, differences in the domain of natural kinds seem to provide the most straightforward counterargument for languages simply mapping onto the external world or onto our conceptual representations of it. The domain of natural kinds is one for which it has been argued that structure in the world will give rise to shared elements of meaning (Berlin, 1992; Hunn, 1977). But again, the broad similarities are accompanied by differences. Primary names that label rich, important, frequently used conceptual content for members of traditional societies sometimes point to highly impoverished representations for members of urban societies (Wolff, Medin, & Pankratz, 1999). Conversely, some significant

[48] As it often happens, to one study showing certain results there is a complementary study revealing different directions of interest within the discipline and offering opposite results (see for example: Youn et al., 2016). Instead of being the confutation one of the other, these studies show how the way of assessing the linguistic meaning of words substantially varies the type of possible observations and outcomes. This should be of no surprise considering not just the difficulties in representing but also in defining the notion of linguistic meaning.

groupings may lack labels in various languages, as in the case of labels for plants versus animals mentioned earlier, as well as groupings within them that are appreciated but of lesser cultural utility (for review cf. Malt, 1995). As Malt et al (2015) argue, the degree to which words of a single language can directly reveal shared concepts for a domain for speakers of different languages will fall on a continuum from poor to better (but still imperfect).

Similarly, we can reject the hypothesis that culturally variable concepts lead to culturally variable meanings. A remark is here in order; we are not rejecting the idea that *some* culturally variable concepts may give rise to variabilities in word meaning. We are nonetheless rejecting the idea that this mechanism may constitute the explanation of the across-the-board relation between thought and culture. In fact, this perspective would allow us to take into account the widespread cross-linguistic diversities while maintaining that concepts are the causal agents: different cultural needs, interests, ecologies, and experiences lead to different concepts that, in turn, give rise to different linguistic contents. However, this answer fails to provide an explanation for many observed cross-linguistic lexical differences. Many differences that have been the focus of recent interest are not readily related to specific cultural needs, interests, ecologies, or experiences. What common aspects of culture would lead Italian and Russian speakers to encode two shades of blue under two different labels, compared to aspects of the English culture where this difference is not available? What different cultural needs, interests, ecologies, or experiences would lead English speakers to make a lexical distinction between spatial relations called *in* and those called *on*, Dutch speakers to further split the relations encoded in English *on* into *op* and *aan*, and Spanish speakers to label all of those by a single word, *en* (Bowerman, 1996a)?

Evan completely reversing the arrow of causality does not seem to provide us with a coherent and comprehensive solution. Making language the only causal agent in concept formation (as it would be in linguistic determinist views) does not allow us to make sense of many evidences from the cognitive science literature. For example, what is not captured in a given language in its lexical contrasts may be salient for non-linguistic reasons, and there is no trace of evidence speaking for the complete incapability of speakers without certain lexical distinctions to perceive certain features not expressed in their languages. For instance, English speakers do not lexically distinguish male from female cousins, as speakers of some other languages do, but they still need to choose gender-appropriate birthday gifts for their cousins, which they seem totally able to do. Furthermore, as we have seen in the introductory chapter, in some domains in which effects of

language on nonlinguistic representations have been found, the exact interpretation of the effects remains up in the air. For instance, the presence and strength of language effects on color perception seem to vary depending on the hemisphere to which the stimulus is presented and the speed at which responding occurs (Regier, Kay, Gilbert, & Ivry, 2010), and some languages have shown gender effects while others have not (Vigliocco, Vinson, Paganelli, & Dworzynski, 2005).

Looser kind of mappings have been argued for. For example, Malt and colleague (2010) propose a constrained by flexible mapping possibility. The authors start from the observation that, at least is some domains, there is some dissociation between how people think about a domain and how language encodes it and that there are indeed constrains on the way languages encode our knowledge of the world. For instance, in a previous study (Malt et al., 1999), the authors showed that in contrast to the cross-linguistic differences in naming patterns of ordinary household storage containers, American English, Argentinean Spanish, and Mandarin Chinese show remarkably consistent similarity judgements. Furthermore, in a recent study, Malt and colleagues (2015) show that it may be possible to discern shared elements of meaning across languages that may indicate constraints on cross-language variability and reflect some common underlying aspects of how knowledge is represented (their case study is on how various languages encode aspects of human locomotion), suggesting there might indeed be a shared conceptual space from which languages draws despite their differences.

In conclusion, it becomes clear that even though we might not be able to decide upfront what kind of link is there connecting language and thought, the two can be dissociated and this dissociation is what allows us to pose the question of their relationship in the first place. Language may be not always the determinant factor in the way people think about things (also given the myriad ways people learn about parts of the world and saliences in the world aside from language). Nonetheless, language differences speak strongly against the language-as-mapping perspective, leaving open the possibility on an influence of language on thought. However, under this hypothesis it is necessary to explain what causes languages to be different in the first place, given that they are not direct reflections of the structure of the worlds or of our mental structure. Besides the constrains mentioned before and the (now obvious) dependence of language on more basic cognitive mechanisms and on the limitations imposed by our bodily-physical constitution, languages show significant elements of arbitrariness in the development of naming patterns. One hypothesis, that can build on the basis of the Saussure's view on linguistic meaning (as introduced

it before), is that language differences may arise through mechanisms independent of conceptual differences and that may allow us to explain how linguistic differences come about. The general conclusion that we can draw from this brief discussion is that language should be thought of as a categorization device, more than as a mere mapping device, operating on principles not reducible to those of other aspects of cognition. The next question follows automatically: what does it mean for language to be a categorization device? On what principles might it operate? In the next section, we will further develop on the insights provided by the Saussurian and structuralist perspective, offering a (linguistic based) view on how languages organize meaning.

2.3.3 Language as a categorization device: the language-as-system view

At the center of the argument, as we have already seen, there is the idea that languages are not strictly bounded to the organization of the world or of the non-linguistic concepts in determining where they draw their categorical boundaries. In the history of linguistic thought, the independency of languages in carving up the continuum of the matter of thoughts and of sounds has been a focal point of the structuralist perspectives. The foundation of this argument lies in the Saussurian notions of *radical arbitrariness,* of *language as form,* and in the notion of the *value* of linguistic units as determined within the linguistic system. This perspective finds its grand development in Hjelmslev's quadripartition of the sphere of the *content* and of the *expression*.

In the Saussurian formulation, languages segment the phonic matter and the conceptual matter projecting limits that arbitrarily determine the identity of the reciprocal space of sounds and meanings. Before being linguistically organized, the matter of the contents and the matter of the possible sounds is amorphous, i.e. it is an indistinct *continuum* when considered from the point of view of language. In this sense, languages are not bound to the organization of a pre-determined thought; the *continuum* of the matter of thought is linguistically amorphous outside of language. Thus, Saussure does not deny that there is a world of perceptions and of concepts outside of language and that it can be studied by other disciplines, such as psychology (De Mauro, 1967a, pp. 439–440). When it comes to understanding the relationship between language and the matter of concepts, the principle of *radical arbitrariness* allows to explain how and why the same experience can be represented in different languages with different numbers of terms and with different relations between the terms within the language system, such that a language unites what

other languages divide or vice versa. This does not imply that languages have no limitations in the ways in which they carve up the world and Saussure seems prone to admit these limitations to the principle of arbitrariness[49]. In reading Saussure's unpublished works, we can see the importance that the study of the limitations of the arbitrariness had for the author; in the Saussurian perspective, it was fundamental to address the cognitive limits and the universal forces that steer languages and, thus, of everything that is before, after, and with the *langue* (De Palo, 2016, p. 26). Language has to comply with the limits imposed by its components (as, for example, the phonic-acoustic matter) and by the (biological and cognitive) nature of its users (De Mauro, 1982a, pp. 11–16).

It seems important to further underline this point given that, on the one side, it specifically addresses the issue of the relation between language and the world in the framework we laid down for our discussion and, on the other side, it prevents us from endorsing some of the most problematic theoretical standpoints of the post-Saussurian structuralism. This idea of the limitations of the arbitrariness of language should not be misunderstood. In fact, it is not that the principle of arbitrariness is limited *per se*; languages are neither logically nor naturally motivated in the way in which they draw their categorical distinction[50]; but, at the same time, languages are not independent from the laws that govern the individual, i.e. they do not exist outside of the realm of human activities and, therefore, are intrinsically bounded to the human biological constitution. Languages are always grounded in the *faculty of language (langage)*, which is the innate, natural, and biological human faculty to constitute and learn a language. Hence, the faculty of language (and consequently the language we speak) is never completely separated from the living organism and the brain of the individual speaker (so much so that Saussure himself endorses the idea that the faculty of language is biologically situated in Broca's area) (De Palo, 2016, p. 43). Saussure

[49] It is widely recognized that the principle of arbitrariness does not imply a total freedom in the way in which languages carve up the matter of sounds and of contents. "When the structuralist says that each language draws a unique set of distinctions in the *continuum* of sound and makes them functional, this does not necessarily imply that there are no general, or even universal, selectional principles governing the phonological structure of languages. Many structuralists, it is true, have expressed the view that the selection of a particular set of phonological distinctions by particular languages is completely arbitrary. But this view, which may be characterized as relativism (in contrast with stronger or weaker versions of universalism [...]) is not essential to structuralism. It is undoubtedly the case that certain phonetic distinctions are more commonly made functional in the languages of the world than are others, as also are certain grammatical and semantic distinctions. Whatever may have been the historical association between relativism and at least certain versions of structuralism, the principles of structural linguistics as they are presented here are compatible with, though they do not imply, at least some kind of universalism." (Lyons, 1977, p. 234)

[50] According to Joseph, the Saussurian view on the arbitrariness of language does not rule out completely the possibility of iconicity in language but the issue is still a matter of ongoing debates (Joseph, 2000, 2015).

does not deny the existence or the importance of a pre-linguistic thought, but he does reject the idea that the pre-linguistic thought could determines the semantic structure of languages. On the contrary, Saussure endorses the view that the semantic organization deposited in the *langue* is important for the articulation of thought once the linguistic competence is acquired. Hence, there is no need to negate the presence of non-linguistic cognitive activities, but we ought to recognize that thought does not have a linguistic organization previous to language and, therefore, language cannot be reduced to a simple nomenclature of pre-existing mental contents (Gargani, 2007).

At the heart of the Saussurian proposal is the idea that language is, within the limitations mentioned above, a *form* - in the sense of a system of distinctions that depends on the relationships between the terms of the system - that imposes itself on a *substance* or *matter* of the contents and of the sounds. The notion of language as a *form* finds its most comprehensive development and understanding in Hjelmslev's proposal. Hjelmslev develops the Saussurian notion of *structure* focusing on two levels of the claim; on the one side, language is a structure in the sense that the linguistic items can be identified only by means of their functions within the structured system of reciprocal relationships (structure as organization). On the other side, Hjelmslev radicalized the notion of language as a *formal system* that underlies the concrete linguistic manifestations (structure as abstraction) (Lepschy, 1968, p. X). These two levels of analysis are strictly interrelated and point towards the possibility (and the necessity) of a structural approach to semantics. The Saussurian distinction between *form* and *substance* - and the specification of language as a *form* - are further broken down by Hjelmslev in the quadripartition of *substance* and *form* of the content and *substance* and *form* of the expression. As in Saussure, the linguistic sign is defined as the relationship between the *signified* and the *signifier*. In Hjelmslev's perspective, each side of the sign is further decomposed in two levels (the level of the *form* and the level of the *substance*). The linguistic sign is established in the relationship between the *form* of the content and the *form* of the expression. On the one hand, we have the *form* and the *substance* of the expression (i.e., of the *signifier*) and, on the other hand, we have the *form* and the *substance* of the content (i.e., of the *signified*). The relationship between the form and the substance (within each level of the sign) is that of a selection in which the substance picks out the form; a substance is by necessity the manifestation of the form, i.e. it is always formed-substance. Hence, given the substance, we can grasp the form that it manifests. The opposite operation (grasping the substance in the observation of the form) is not possible, since the form can manifest itself in any substance.

Hence, the distinction between *substance* and *purport* (not-formed matter) becomes relevant to understand the nature of linguistic meaning. The *substance* is, at any given time, the manifestation of a form, whereas the same *purport* (for example the graphic matter) can manifest different forms (as, for example, written English or written Italian). In this theoretical framework, we can also understand Hjelmslev's critique to the Saussurian thesis of the pre-linguistic nebulosity of the content found in the *Prolegomena to a Theory of Language* (1961). From the point of view of language, we never actually experience a possible amorphous content of thought, since it always appears to us as already linguistically formed-substance (De Mauro, 1967a, n. 225). By means of the study of language, and especially by means of within-language and cross-linguistic observations of linguistic signs and expressions,

> we thus see that unformed purport extractable from all these linguistic chains is formed differently in each language. Each language lays down its own boundaries with the amorphous "thought-mass" and stresses different factors in it in different arrangements, puts the center of gravity in different places and gives them different emphases. […] Just as the same sand can be put into different molds, and the same cloud take ever new shapes, so also the same purport is formed or structured differently in different languages. […] Purport remains, each time, substance for a new form, and has no possible existence except through being substance for one form or another. We thus recognize in the linguistic *content,* in its process, a specific *form*, the *content-form*, which is independent of, and stands in arbitrary relation to, the *purport*, and forms it into a *content-substance*. (Hjelmslev, 1961, pp. 51–52)

The notion of radical arbitrariness and the radicalization of the principle of language as a *form* become a theoretical device that allows to understand the diversities we observe in the semantic structure of different languages. The matter of the expression, as for example the *continuum* of the vocalic sounds, becomes differently-formed substance in different languages: for example, languages may distinguish between three or five or seven vowels. Similarly, the same matter of content, as for example the continuum of the color spectrum, becomes differently-formed substance in different languages: for a given area of the spectrum, English distinguishes between four colors (*green, blue, gray, grown*) whereas Welsh distinguishes between three colors (*gwyrdd, glas, llwyd*) (cf. Hjelmslev, 1961). In this framework, the way towards the outline of a possible

structural semantics becomes evident.[51] Semantics, and lexical semantics in particular, has to be understood in the framework of a broader notion of language as system of values, i.e. as an autonomous system of interdependent terms. In this perspective, the study of semantics pertains to the domain of the substance that reflects the system of dependencies internal to the form of the content (cf. the Note to "Per una semantica strutturale" by M. Prampolini in Hjelmslev, 1981, p. 147).

In the previous paragraphs - and in line with distinguished interpreters of the Saussurian thought, we have already attempted to question the commonly shared assumption that, according to the Saussurian perspective, the notion of *langue* would account for the whole of the classificatory and categorizing roles of languages while neglecting the centrality of the language and thought relationship. According to De Mauro's reading of some influential passages in the Saussurian work, we have already mentioned how Saussure is interested in highlighting that the thought is linguistically amorphous previous to language, and not in ruling out the possibility of a world of perceptions and ideas independent from language. Furthermore, we have already referred to the contributions of the cognitive sciences and of developmental psychology in showing the richness of the pre-linguistic thought and the importance of perceptually grounded categorization for semantic as well as conceptual organization of adult human thought. In reiterating these observations, we ought to remark that, even in the Saussurian perspective, the matter of the content is not psychologically completely unformed but rather it has grooves that stem from our perceptual and psychic constitution. The formativeness of the linguistic sign is always interdependent with the pre and extra-linguistic reality. Starting from the distinctions made available from our constitution as humans, languages build and produce points of view, i.e. they determine forms of organization of meaning (cf. De Palo, 2016, p. 69). In this perspective,

> La lezione dell'autonomia del significato, più che indicare un completo rigetto da parte di Saussure della componente psicologica dei fatti linguistici, porta a compimento la svolta filosofico-linguistica humboldtiana attraverso l'indagine del- la diversità delle lingue e della diversità dei comportamenti linguistici individuali. In questa esplorazione, la linguistica ci restituisce la mappa di un territorio eterogeneo in cui lingua e soggetto, forma e sostanza, langue e parole, disegnano le linee di ricerca di una

[51] In the next section of the chapter, we will briefly outline the structural semantic approach put forward by Hjelmslev when discussing the structure of the lexicon and, more precisely, the componential views of meaning.

> teoria della significazione in cui la langue è la bussola del linguista (De Palo, 2016, p. 48).

Thus, it is within the framework of the Saussurian linguistics that we find the need to encompass the individual, both in terms of biological components as well as of social factors, to understand the categorizing role of language. On the one side, as we have pointed out many times, the *value* of the linguistic units – the meaning – can be established only by means of the differences and opposition with other elements in the system. Taken at the extreme, the Saussurian perspective suggests that there are no positive elements in the way linguistic meaning is established (and used). When analyzing linguistic meaning, the possibility of establishing positive definitions of linguistic units is essentially an illusion, given that the meaning of the units cannot be grasped outside of the relations it establishes with the rest of the system. This is evident in the way we use words.

> Vediamo allora che non è per niente l'idea POSITIVA contenuta in *supplizio* e *martirio*, ma è il fatto NEGATIVO della loro opposizione che fissa tutta la serie dei loro impieghi, permettendo qualunque impiego purché non invada il dominio vicino (Saussure, trad. ita, 2005, p.90).

> Una parola non esiste veramente, da qualunque punto di vista ci si collochi, che grazie alla sanzione che riceve di momento in momento da parte di quelli che la impiegano. È questo che la rende differente da una successione di suoni e che la fa differire da un'altra parola, fosse pure questa composta dalla stessa successione di suoni. Poiché non vi è alcuna *unità* (di qualunque ordine e qualunque natura la si immagini) che poggi su altra cosa che su *differenze*, in realtà l'unità è sempre immaginaria, la differenza sola esiste (Saussure, trad. ita, 2005, p.94).

Hence, as in the formulation used by Saussure in the *Course*, the value of a unit is determined *par ce qui l'entoure* – by the elements that surround it. This formulation is fundamental in at least two passages of the *Course* (that we report here in Harris's English translation):

> The French word mouton may have the same meaning as the English word sheep; but it does not have the same value. There are various reasons for this, but in particular the fact that the English word for the meat of this animal, as prepared and served for a meal, is not sheep but mutton. The difference in value between sheep and mouton hinges on the fact that in English there is also another word mutton for the meat, whereas mouton in French covers both." In a given language, all the words which express neighbouring ideas help define one another's meaning. Each of a set of synonyms like *redouter* ('to dread'), *craindre* ('to fear'), *avoir peur* ('to be afraid') has its particular value only because they stand in contrast with one another. If *redouter* ('to dread') did not exist,

> its content would be shared out among its competitors. [...] So the value of any given word is determined by what other words there are in that particular area of the vocabulary (Saussure, 1983, pp. 236–237).[52]

> A linguistic entity is not ultimately defined until it is delimited, i.e. separated from whatever there may be on either side of it in a sequence of sounds. It is these delimited entities or units which contrast with one another in the mechanism of the language (Saussure, 1983, p. 215).

The value of a linguistic unit is given from its relationship with the words that surround it 1) among competitors and 2) in a sequence of sounds. We will see how this key formulation can be reinterpreted in the light of distributional accounts of linguistic meaning. The meaning of linguistic units is relational, and it is grounded in the uses we make of language in the written and spoken chain. To help us better understand this point, allow us some more specific remarks about the Saussurian perspective. The relational and differential view of language needs to be integrated within a broader perspective.

The system of language (*langue*) and the more general principles of how language works can be fully understood only in connection with the "third principle" of the Saussurian linguistics, as theorized and understood in De Mauro's *Introduzione alla semantica* (1965), which we have already introduced before. The functioning of language cannot be understood outside of the social context (De Mauro, 1965, pp. 152 et seq. and 169 et seq). It is within the social practice, in the *masse parlante (speaking mass),* and in the development of the language in time, that we standardize and actualize the distinctions of the linguistic system. As we have seen, the *langue* as a system is not independent from the individual and idiosyncratic acts of *parole*; on the contrary, it is in the dialectic relationships between *langue* and *parole* that we are actually able to understand how the language system can balance between the individual and idiosyncratic uses of words and the need of efficient communication that substantially implies the intersubjectivity of language. The abstract classes that compose the *langue* are stable but not logically closed categories. Words do not have stable and fixed margins of meaning; rather, they are characterized by an intrinsic plasticity that allows the extension of these abstract classes of signification to new uses and senses both in the process of communication between individuals as well as in the development of

[52] For citing Saussure, we refer to the first edition of the *Course in General Linguistics* translated by Roy Harris in 1983.

languages over time. Therefore, the meaning of linguistic signs is *indetermined*, i.e. it is open to the fluctuations of the senses that we observe in the linguistic manifestations. The *indeterminacy* of meaning describes the very mechanism by which language operates as classificatory device; we can characterize the language system as a web of semantic criteria of pertinence/significance by means of which we are able to link an individual signification to one or to another class of significations.

Through this discussion of how languages carve up reality – on how we can understand the relationship between language and the world - we seem to have reached some (at least provisory) conclusions: 1) languages are not completely bound to the external organization of reality but there are limitations to the arbitrariness of language; 2) linguistic categories should be understood as web of pertinences and not as fixed and logically closed containers; 3) linguistic meaning (in looking at it from both the perspective of the *langue* and of the *parole)* does not exist in isolation from the speakers, i.e. it exists only by means of the linguistic practice within the community of speakers. We have tried to show how the Saussurian perspective leads the way to a reconsideration of the role of the speaker in both its biological, mental as well as social components. In the next and conclusive section, we will further develop this line of thought, showing how it is in the processes of meaning creation - of "doing meaning" - that we can start grasping the formative role of language, grounding our reflection of the specificities exhibited by the language as a system.

2.3.4. The dynamicity of "doing meaning"

As we have previously observed, languages, and hence linguistic meaning, do not and cannot live in isolation from the community that speaks them. As De Mauro points out, words do not have an intrinsic semantic virtue if we isolate them from the speakers. Words do not grant by themselves the transmissibility of a univocal content; they acquire this ability only in relationship to the uses we make of them (De Mauro, 1965, p. 170). Therefore, if our aim is to outline a possible way of understanding linguistic meaning, we cannot avoid taking the perspective of the speakers, i.e. of the individuals that experience the world and make use of language.

The way in which we construct meaning at the level of the individual cannot be separated from the experience we have of the world and from the constitution of our perceptual and cognitive system. Hence, we can now try to close the circle that took us from cognitive perspectives on the

thoughts-words relationships to the Saussurian notion of meaning and of language as a system, as well as to set the ground for the proposal that will follow. The relational-differential view of meaning and a cognitive perspective on meaning are not two fully incompatible views. The general view put forwards by cognitive semantics leads towards highlighting the non-autonomy of linguistic meaning; a theory of semantics has to be understood as a theory of how we comprehend linguistic expressions (Marconi, 1992, p. 431; Violi, 1997, p. 45) and it should point towards the description of linguistic meaning as cognitive content. In order to understand what linguistic meaning is from a cognitive point of view we need to take into account the realm of what is non-linguistic, starting from the cognitive processes and embracing the contents we acquire from experience. From the point of view of the individual that experiences the world, linguistic meaning is indistinguishable from the rest of our experience which is, at the same time, perceptual, physical, psychological, mental, cultural and social (Violi, 1997, p. 50). This means that there is – most probably - only one level of mental representations in which perceptual, sensory-motor and linguistic information converge.

Consequently, cognitive semantic theories call into question the autonomy of linguistic meaning (a perspective that has often been ascribed to structural accounts of meaning). As we have already pointed out in the previous section, the Saussurian view of meaning is not insensitive to the centrality of the speaker. The web of pertinencies that constitutes the system of language constantly supports and renovates itself starting from the use that speakers make of language, in which the linguistic and the encyclopedic contents of meaning are indistinguishable. It is by means of the fuzziness and indeterminacy of the meaning of words that we can recompose the structural and individual components of meaning; at the center of the analysis, we ought to consider

> […] non tanto il segno preso isolatamente ma, per l'appunto, il 'modo di vita' – potremmo dire – del segno, in necessario collegamento con gli utenti, con gli esseri umani che si muovono nella realtà che li circonda, categorizzano oggetti, esperienze ed emozioni, insomma costituiscono pratiche ma anche saperi depositati, esperienze e culture. Non si assume più, insomma, che vi sia un livello autonomo di organizzazione semantica separato dalle più generali capacità concettuali degli esseri umani, e, analogamente, i fenomeni semantici non sono più studiati come dati puramente linguistici, ma sono messi in relazione anche con il loro sfondo contestuale, il quale è sia di tipo psicologico-percettivo che socio-culturale e pragmatico. Insomma, se noi adottiamo una prospettiva di non-autonomia del linguaggio e della semantica rispetto alla realtà e alle nostre modalità percettivo-cognitive, allora la separazione tra le

> conoscenze della lingua e quelle del mondo, tra i significati linguistici e l'insieme delle conoscenze enciclopediche dei parlanti non ha ragione di esistere. […] non solo le conoscenze linguistiche, ma pure quelle enciclopediche, quelle socio-culturali ecc. Costituiscono quella conoscenza -per così dire - di sfondo che, anche in virtù del suo essere incorporata nei significati delle singole parole, rende possibile e garantisce la reciproca comunicazione e comprensione tra i parlanti di una comunità linguistica (Basile, 2001, p. ?).

Words and their significations are inextricably connected to human experience, inasmuch that "the meaning of language is inseparable from the meaning that the world has for us" ([our translation] Violi, 1997, p. 5). The signification, the processes of "doing meaning", the concrete and individual actualizations of the *signified*, cannot be separated from the activities by means of which we effectuate them. The value of a linguistic unit does not point towards a fixed meaning, but rather towards a complex *semantic potential* that, on the one side, is determined by its various uses in context and that, at the same time, creates its own contexts of uses (Basile, 2012, p. 21). It is again in the notions of "sense" and "acceptation" that we find a useful interpretative key. In the concrete acts of the linguistic activity - in the communicative practice - we are exposed to the actual semantic values of the linguistic signs; the "sense" is an individual phenomenon connected to the various experiences and knowledge of the subject, which becomes pertinent and identifiable from a linguistic point of view in the web of similarities and differences of the *langue* that organizes our experiences in a web of language-internal relationships. Conversely, this web of pertinences is constituted in relation to the experiences of the speakers in their actualization and use of language. To understand how Saussure's semantics can be integrated with cognitive accounts of meaning we need focus on the dynamicity of the relationship between *langue* and *parole*:

> Le forme del contenuto, cioè i *signifiés* dei segni, appartengono alla lingua. Nella *parole* individuale non troviamo i *signifiés*, ma le realizzazioni e concretizzazioni dei *signifiés*, cioè (come alcuni ritengono) ciò che Saussure chiama *significations* o *sens*. Per una lingua data ciascun *signifiés* ha un certo grado di stabilità. [...] Per capire la *signification* di un concreto enunciato o testo dobbiamo certamente conoscere la lingua in cui viene formulato ma al contempo dobbiamo valutare le circostanze di *parole*. [...] La significazione ha dunque un ruolo attivo nel configurarsi delle forme sintattiche degli enunciati ed è questo uno dei punti attraverso cui, nella visione di Saussure (qui oserei permettermi di esprimere una piena adesione), l'uso individuale produttivo e (si noti) ricettivo non si limita a realizzare ed eseguire, ma sollecita dinamicamente, corrode o ridetermina il sistema di forme della langue (De Mauro, 1994, pp. 124–125).

Hence, the process of "doing meaning" is a dynamical interaction between the values of the words as they become standardized by the practice of the linguistic community in the structure of the language as a system and the individual acts of *parole*. In the acquisition of linguistic meaning, we learn both the way in which our native language refers to the world of the objects and of the state of affairs as well as the habits, knowledge, and representations of the world deposited in and conveyed by language (Basile, 2010a, p. 28). The cognitive as well as the communicative functions of language are actualized *in* and *by means of* language as a tool of cognitive and social mediation of our relationship with the world and with our social interactions (Basile, 2012, p. 42). When focusing of the mental level of the individual speaker, the question about the relationship between language and the world, or between the structure of language and the structure of the world, loses its relevance; languages are not something separated from the rest of the world, but they are rather part of the very same *milieu* (Basile, 2005, p. 72). Starting from this considerations, Basile outlines the need for a "global" semantic approach that would allow us to grasp all aspects connected to the dynamical process of "doing meaning", considering the speaker in its relationship with other speakers and with the surrounding world (Basile, 2001, p. 32, 2012, p. 44).

In conclusion to this concise discussion about the dynamicity of linguistic meaning and of how we "do" meaning, we want to draw the attention to De Mauro's insightful observations in the *Epilogo in terra* (epilogue on earth) in the conclusions to *Minisemantica* (1982a). By means of a brief, clear, and enlightening argument, De Mauro delineates the theoretical space for a possible resolution of the long-standing fierce dispute between universalistic and relativistic perspectives on language and culture. On the one side, universalistic views of language fail to provide a substantial explanation to language diversities while leaving open the possibility of cross-cultural and cross-linguistic communication. On the other side, language deterministic views defend the legitimacy of language diversity while sacrificing the inter-cultural and inter-linguistic communicability. De Mauro cuts down the problem to its bone: both approaches fail to understand the very nature of language because they view language as a static, given-once-for-all instrument that is bounded by a one-to-one relationship to the cultural and mental world (likewise understood as static and given-once-for-all worlds) (De Mauro, 1982a, pp. 157–158). But, as De Mauro states, this is not actually the case of our experience with language. Instead,

> le lingue sono codici non interamente calcolabili, codici creative, e ciò sul versante semantico comporta una permanente possibilità di estensione dei significati a nuove

> accezioni e sensi e si accrescimento del vocabolario. […] Le lingue sono diverse perché si diversificano di continuo ciascuna al suo interno e in rapporto alla sua tradizione. E grazie a tale flessibilità indefinite, che le nostre nozioni di "creatività" e "indeterminatezza semantica" hanno cercato di identificare e comprendere, esse possono portare fino a quel caso limite del diversificarsi indefinito che è l'istituzione di reciproche analogie, di somiglianze e scambi. Nessun problema, dunque, ad accettare nello stesso tempo l'evidenza della diversità profonda delle lingue e quella della possibilità di mutua comprensione tra genti di lingua diversa. (De Mauro, 1982a, p. 159)

De Mauro's words seal from a theoretical perspective the attempt made in this chapter to reconsider the current debate in the cognitive sciences about the relationship between language and thought and between language and the world in the light of the notion of language as a dynamic process and as a dynamic product.

In conclusion, we have seen how we can come to a multifaceted definition of linguistic meaning that constitutively involves every aspects of our reality. On the one side, the meaning of words appears as a stable but not logically closed set of malleable categories that depend on their existence within a community of speaker; on the other side, linguistic meaning exists only in and by means of the concrete and individual uses of the words and, as we have seen, it is fundamentally linked to the perceptual-psychological ways in which we experience and categorize our external as well as internal reality. Hence, the study of meaning is always placed in a precarious balance between the (relative) stability of the *signified* and the infinite possibilities of the concrete acceptations and senses.

We gain three main insights from the perspective outlined in this chapter as far as the language and thought relationship is concerned: in the first place, meaning in language is not the same as the content of our mental representations (however we define them). Language exhibits its own specificity and its own perspective on the world. In the second place, this separation between language and thought can hardly be identified at the cognitive level, i.e. we cannot simply hypothesize two sets of mental representations providing us with separate contents. Language is both a cognitive phenomenon, and hence it depends on to both our domain-general processes as well as on the non-linguistic knowledge we have of the world. If there is just one level of mental representations (here, we have been approaching the issue from the point of view of theoretical linguistics, but we will see that researches in the cognitive sciences lead to similar conclusions) and if linguistic meaning is for the speaking subject a whole with the rest of his or her knowledge of the world, how can we go about the separation between language and thought? We come to the

third and final point of our discussion: language, and more specifically the language system, arises from the coordination of speakers within the communicative practice. Hence, language is not just a cognitive phenomenon, but it is, at the same time, a socially shared phenomenon. And it is precisely in the combination of these two aspects that we can understand the nature of linguistic meaning. We can start to look for a language-specific level of meaning in the systematicity of language uses, which gives rise to those abstract classes (the *signifiés*) in an organized web of intra-linguistic relationships. It is in the language as a system of interrelated and inter-determining units that we should start to look for the specificity of meaning in language.

In the next chapters we will address well-consolidated theories on the structure of language (and more specifically of the lexicon of the language) and we will try to outline what main aspects of meaning they allow to grasp. In the third chapter, we will mostly follow a classical historical reconstruction of the field. We will inevitably come back to the language-as-mapping hypothesis and how it reshaped the study of linguistic meaning and cognition more broadly. In the fourth chapter, we will turn to a more specific issue, i.e. how language needs to be structured in order to be processed for production and comprehension. More broadly, the chapter addresses the issue of how language is to be understood if we look at it as a cognitive phenomenon. The fifth chapter turns to the other side of the issue, i.e. how language should be understood if we look at it from the side of its systematicity in the uses we make of it. In chapter six, we will try to recompose the two approaches, outlining a way to look at how language becomes a constitutive ingredient of how we create meaning.

Chapter 3

The structure of the lexicon: from linguistics to psychology

3.1. Structure in language and structure in cognition: outline of the chapter

According to what argued in the previous chapters, the observation that language (viewed both as a mental as well as a linguistic fact) shows a systematic organization of its contents should not sound like a curious incident. If we ask an English native speaker the first words that come to his or her mind when we say, for example, the word *brother*, we will easily note that there is a high recurrence of the same associations; when we think about the word *brother* we also seem to think about the word *sister*, *sibling*, *family*, *father*, *mother*. Is this just a matter of chance or is it that the regularities exhibited by word associations tell us something more about how we organize knowledge in our minds? Why do we associate the word *platypus* with the word *animal*, *tall*, *Australia*, *water*, but also *weird*, *strange,* and *unique*? Does this have anything to do with language? Or, more specifically, do these regularities depend exclusively on our experience of the world or do they have anything to do with the meaning of words, i.e. with how language organizes knowledge and experience?

From the mistakes people make while speaking as, for example, when they select a word instead of another (the so-called *slips of the tongue*), we can observe various interesting patterns by means of which words are associated in our minds; in "*he told a funny antidote*" the actual word would be *anecdote* and in "*the inhabitants of the car where unhurt*" we probably meant that the *occupants* of the car were unhurt (examples from Aitchison, 1987, p. 10). Why does it happen that we say *antidote* instead of *anecdote*, or *occupants* instead of *inhabitants*? There is clearly some kind of relation between the words that are wrongly selected and the word we probably wanted to select. In some cases, words can be wrongly selected because they have similar sounds. In other cases, slip of the tongue errors exhibit similarity in content. Why does language exhibit such an organization? On the basis of what principles in language organized? Is there something inherently linguistic about the way in which language is organized?

These and similar questions have been the focus of much of the research paradigms interested in the problem of the structure of our experience and of language in mind. The importance of free words association tasks dates back to the late '800 and it is still a fundamental method in psychological approaches interested in exploring the relationship between language and thought. We will extensively discuss some of the major theories concerning the organization of words in the mind and some of the major issues with free word associations in the next chapter.

Here, we want to start for a slightly different perspective: our aim in this chapter is to briefly review some of the main theories that have been concerned with the structure of language and with the nature of linguistic categories. Some preliminary observations are here in order.

In the first place, as we have showed in the previous chapters, we have ruled out the possibility of a strictly autonomous view of linguistic meaning by focusing on the dynamic constitution and understanding of meaning that could account for both linguistic and non-linguistic facts. However, the possibility of separating linguistic from non-linguistic content (meaning) is one of the main themes of this chapter: the approaches we will review here divide into approaches that endorse an autonomous view of language (i.e. structuralist approaches) and approaches that reject such autonomy and highlight the encyclopedic nature of linguistic meaning (i.e. cognitive linguistic/semantic approaches). As we have already pointed out, we do not endorse an autonomous view of linguistic meaning in the way in which it will be intended in this chapter. However, starting from structuralist approaches allows us to retrace the history of the debate about how language may be organized at a linguistic level. In the second part of the chapter, we will turn to more specifically cognitive issues and we will show how both the notion of linguistic category as well as the notion of system shifted after the cognitive revolution(s).

In the second place, we will try to shape our discussion around another issue that has been fundamental in the previous chapters: the relationship between *langue* and *parole*. We will try to highlight how the various theories have given importance to either one or both poles, and we will outline some of the most outstanding issues that arise in those proposals that take the opposition between *langue* and *parole* as dichotomic in nature. We will also see how the distinction between *langue* and *parole* was progressively abandoned in cognitive approaches to meaning. In the conclusion to the chapter, we will argue for the necessity of reconsidering the role of the *langue* (as outlined in 2.3.3.) and of reintegrating it in cognitive approaches to language.

In the third place, this chapter should be viewed as mostly an historical reconstruction of theories on the structure of language, and not as much as an integral part of our proposal. We will retrace the historical and theoretical paths taken by researches in the study of the structure of the lexicon and we will try to outline the strides that connect linguistic and psychological accounts of the systematicity of language. In order to provide a (at least partially) linear discussion that could encompass the multifaceted set of problems, questions, and arguments that implicitly and explicitly contributed to the investigations on the structure of language, we will mostly embrace a historical perspective and we will try to focus on the major shifts in the questions that underlie the various accounts and theories of the last few centuries.

We will start with the analysis of those proposals concerned with the structure of language *in language*: at the core of the structuralist argument, as we will see, there is the need to develop and to establish methods that could explain language and linguistic phenomena from the point of view of the autonomy of meaning. We will set out to discuss the Saussurian perspective, which will give us the lead for the analysis of structuralist approaches to meaning. Within the framework of structural semantic approaches, we will outline two main directions of research: in the first place, we will focus on the notion of semantic field from its origins as an instrument to describe and understand language changes over time, to its application in more recent paradigms of structuralist research. In the second place, we will try to grasp the theoretical needs behind the rise of compositional semantic approaches. In conclusion, we will present some interesting work at the intersection of semantic field and compositional approaches. Throughout the discussion, our endeavor will be to show both the positive outcomes of such theories as well as some of their main limitations, especially in the light of the theories that will follow thereafter.

In the second part of the chapter, we will see how the shift from an autonomous to a non-autonomous view of meaning (promoted with the rise of the cognitive sciences) will reframe some of the main problems under investigation. In the first place, we will review attempts to reconnect the more linguistic side of meaning with the broader knowledge we have about the world as well as, more broadly, the study of meaning in language with the study of meaning in the mind. We will also try to show how this shift substantially led to the language-as-mapping view (cf. 2.1.), and, hence, to the debate on the relationship between conceptualization and lexicalization we have encountered in the previous chapter. These questions led to interesting developments and novelties in the study of the structure of language and of the lexicon. On the one side, theories of prototypes

introduced a new way of looking at the nature of conceptual (and therefore linguistic) categorizations that jeopardized the traditional notion of category and its applicability and efficiency in the study of concepts and in the description of linguistic categories. Similarly, the integration of encyclopedic knowledge in the study of linguistic meaning brought to reconsider the notion of semantic field and, especially, to delineate a theoretical structure that could encompass both linguistic meaning as well as other types of knowledge. An example of this last enterprise is the rise of *frame semantics*.

It should be remarked upfront that the aim of this chapter in not to rewrite the history of lexical semantics, but rather to offer an overview of the development of research paradigms interested in the structure of the lexicon and, in part, on the structure of linguistic categories. We will try to outline the main questions underlying the approaches taken into consideration in the following pages and we will discuss their theoretical and methodological outcomes (see figure 7 for a summary of approaches).

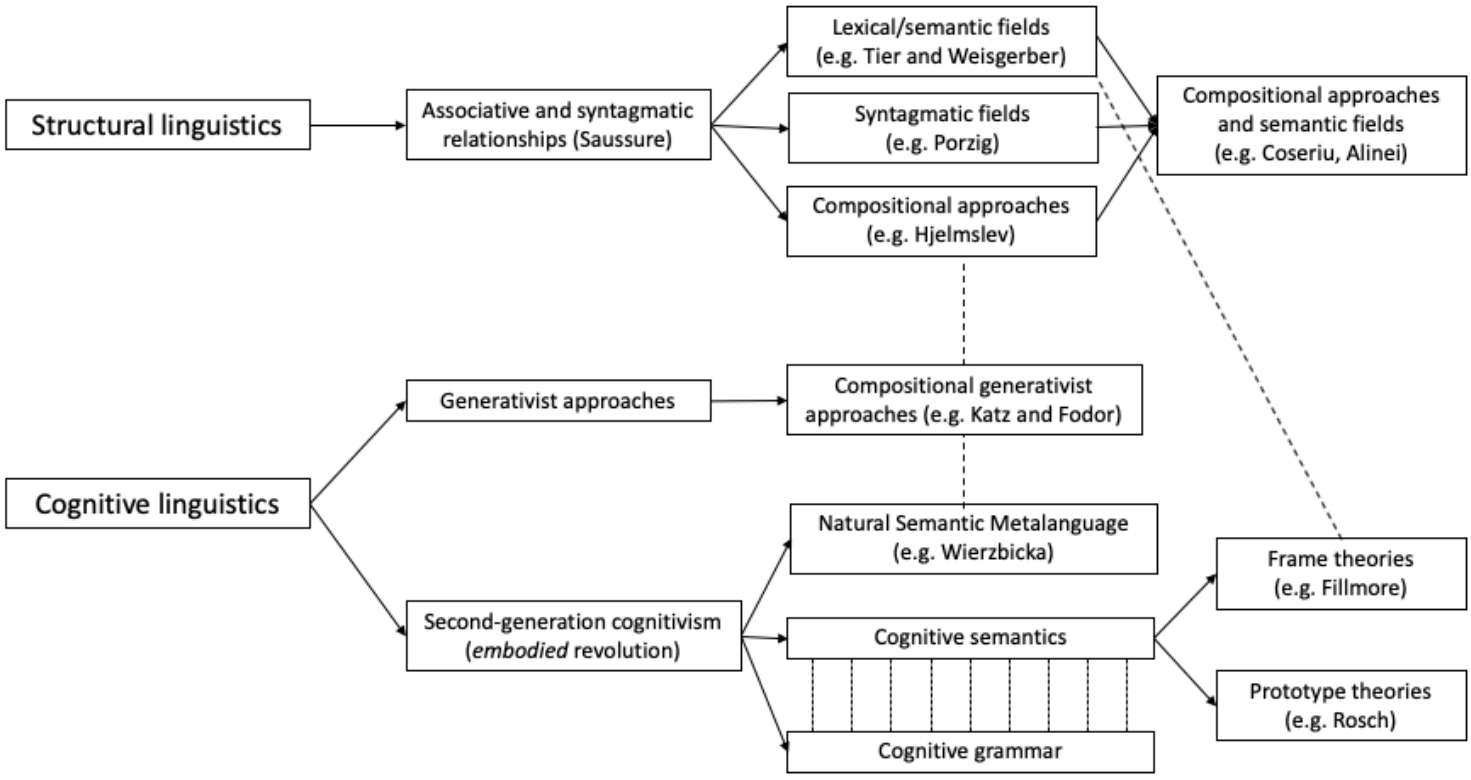


Figure 7. Summary of approaches we will be discussing in the chapter organized by the filiation from the two general approaches we are considering, namely structural semantics and cognitive semantics. Full arrows represent direct filiation. Dotted lines represent connections with other theories. Lines between cognitive semantics and cognitive grammar indicate that the two approaches, although tackling two different issues in the study of meaning, are inherently connected in a unitary framework.

3.2. Structuralist approaches to the structure of language: relationships between words

3.2.1. The theoretical kickoff: associative (paradigmatic) and syntagmatic relationships

Let us begin with the view from linguistics; how are words organized in language? Once again, our starting point is Ferdinand De Saussure. The centrality of Saussure's work for the study of semantics has long been underestimated and it is only in recent years that his ideas became food for thoughts. This is especially true for those semantic theories originating in the North-American intellectual context, such as those we have been reviewing so far, whereas the European structural semantic approaches often cite Saussure as one of their most relevant sources. However, almost as often, they led to observations and conclusions that could hardly fit in the original Saussurian framework.

The reasons behind this lack of interest for the Saussurian thought in the study of semantics are connected to the way in which the original version of the *Cours de linguistique genérale* was edited by Charles Bally and Albert Sechehaye, who omitted most of the more intrinsically semantic notes of Saussure. Consequently, this led to the idea that Saussure himself was not interested in study of semantics at all, and gave rise to the suggestion and the possibility of studying language without studying meaning. At the same time, formal theories interested in establishing a logical theory of meaning were incompatible with Saussure's semantic ideas that, if correctly interpreted, mined the very roots of any understanding of languages as artificial symbolic systems (De Mauro, 1994, pp. 119–126). Conversely, De Mauro's reading of Saussure's work emphasizes, on the one hand, the importance of semantics in the development of Saussure's linguistic theory and, on the other hand, he focuses on the centrality of Saussure's semantic ideas for an adequate theory of the meaning (De Mauro, 1994, p. 121). We have already seen how the notions of *form* and *substance* and the idea of the systematic nature of the linguistic *value*, combined with the dissolution of the dichotomy between *langue* and *parole* in the "third principle" of the Saussurian linguistics (cf. 2.2.3 and 2.3.3.), determine the need for a dynamic understanding of meaning that encompasses both the individual as well as the socially shared aspects of language. Furthermore, in Saussure we also find one of the first accounts of how language, and specifically linguistic meaning, might be organized. The notions of *syntagmatic* and *associative* relations are the principles according to which linguistic meaning is systematically organized. As it will become clear later, in the

formulation of these principles Saussure does not seem to renounce the need of integrating the level of the *langue* with the level of the acts of *parole*, the level of the system with the level of the individual and mental.

As we have previously mentioned, Saussure has established the relational nature of linguistic meaning by means of the notion of *value*: in order to grasp the meaning of a word, we need to compare it with linguistic units of similar value, i.e. with those units of the system that are in relation to it.

> In a linguistic state, then, everything depends on relations. How do they work?
> The relations and differences between linguistic items fall into two quite distinct kinds, each giving rise to a separate order of values. The opposition between these two orders brings out the specific character of each. They correspond to two different forms of mental activity, both indispensable to the workings of a language (Saussure, 1983, p. 248).

Thus, two forms of mental activity give rise to two distinct kinds of relationships within language: *associative* and *syntagmatic* relations. Even before turning to the actual discussion of the nature and realm of application of the two kinds of relations, we ought to pause on the fact that, in order to introduce the discussion on the structure of language, Saussure calls into question some form of mental activity. The way in which Saussure faces the question about the structure of language creaks the pure oppositional and relational setup outlined before (Basile, 2001, p. 74) or, at least, the complete autonomy of the linguistic from the psychological aspects of language as envisaged, on the contrary, by post-Saussurian structuralist approaches.

In the first place, Saussure hypothesizes a psychological *faculte d'association e de coordination* (faculty of association and coordination) that is at the base of the human faculty of articulating the phonic and conceptual matter (Basile, 2001, p. 74), i.e. it is at the base of the constitution of the linguistic sign, as well as at the base of the association between units of the system (De Palo, 2016, p. 63). Such faculty manifests itself in the formation of "groups"[53] of words that exhibit particular kinds of relationships. One remark is here in order; it is not clear how we could interpret Saussure's *faculte d'association e de coordination* (or the more specific mental activities that should give rise to systemic relationships) in terms of recent advances in the cognitive sciences. In it now widely recognized that language is grounded and develops on

[53] Cf. De Mauro (1967a n. 246) and Basile (2001, p. 74 note 27) for the notion of "groups" of words.

domain-general cognitive principles, such as categorization, generalization, chunking and so on (e.g. Bybee, 2010), which may be suitable for translating Saussure's intuition into contemporary cognitive mechanisms. Specific research should address this issue and the consequences it might give rise to both in the actualization of Saussure's thought as well as taking Saussure's view as an inspiration for current research in the cognitive sciences. What is of interest here is that, at least to this point, Saussure's thought is not incompatible with current directions in the study of cognition, but it rather suggests that language and linguistic meaning is ultimately grounded into our cognitive abilities. We will leave this point momentarily aside, inasmuch as more research is needed to evaluate the plausibility of Saussure's claims to this regard. We will here offer a brief discussion of the two kinds of relationships that emerge and structure language.

One kind of relationship is introduced by Saussure as follows:

> words as used in discourse, strung together one after another, enter into relations based on the linear character of languages […]. Linearity precludes the possibility of uttering two words simultaneously. They must be arranged consecutively in spoken sequence.[54] Combinations based on sequentiality may be called *syntagmas* (Saussure, 1983, pp. 248–249).

Syntagmatic relationships are *in presentia* relationships. At a syntagmatic level, words get their meaning in relation to what precedes and follows them in the spoken or written chain. This kind of relationship is easily observable at the level of the sequence of the discourse, in which the principle of linearity of the *signifier* is expressed. In the pages dedicated to the presentation of the *syntagmatic* relationship (as in the passage we have cited above), Saussure seems uncertain about whether *syntagmatic* relationships belong entirely to the realm of the *parole* or if they are as well part of the *langue*. On the one side, extended combinations of syntagms show great variation in the way in which their constitutive elements are collocated; especially in sentences, the collocation of the elements seems to depend upon the free choice of the individual and, therefore, *syntagmatic* relationships seem to belong to the realm of the *parole*. On the other side, many syntagms belong to the mnemonic inventory of combinations standardized in the *langue*. And even when new and unknown syntagms are produces, the "type" of syntagma or of sentence belongs to the *langue*. In this sense, sentences and syntagms belong to the *parole* for all that depends on individual choices

[54] Here Harris translates the term *parole* with *spoken sequence*. The term is left in the original version in De Mauro's translation of the *Cours de linguistique genérale* (Saussure, 1967, p. 149).

but, at the same time, they are part of the *langue* in the sense that they are not entirely idiosyncratic (De Mauro, 1967a n. 251). In this uncertainty, we become again aware of the impossibility of fully separating the level of the *langue* and of the *parole* in the analysis of the system and of the relationships that words entail with each other.

The other kind of relationship identified by Saussure is the *associative* relationship;

> outside the context of discourse, words having something in common are associated together in the memory. In this way they form groups, the members of which may be related in various ways (Saussure, 1983, p. 249).

As it seems clear starting from this definition, the notion of *associative* relationship[55] seems quite hard to define. In the first place, Saussure explicitly states that

> this kind of connection between words is of quite a different order. It is not based on linear sequence[56]. It is a connection in the brain. Such connections are part of that accumulated store which is the form the language takes in an individual's brain[57] (Saussure, 1983, p. 249).

Associative relationships are *in absentia* relationships; i.e. they cannot be observed in the "extension" of language (in the written or spoken sequence) because they are based on relationships of inter-replaceability between words that are associated in mnemonic virtual series by means of various kinds of associations (Basile, 2001, p. 76). It is the mind (*esprit*) that perceives the various relationships by means of which words are associated. In his definition of *associative* relationships, we can see how Saussure (once again) appeals to the psychological dimension of the speaker; in order to understand the content of language we cannot completely avoid the individual level of comprehension that is behind each word and that comprises both linguistic as well as perceptual and, more generally, experiential associations. Violi pinpoints the intrinsic

[55] *Associative* relationships are generally called *paradigmatic relationships*; Saussure generally proposes examples to clarify his idea of *associative* drawing from flexion paradigms (De Mauro, 1967a n. 248) and Hjelmslev standardized the use of such term (Hjelmslev, 1961).

[56] Here, Harris is translating the French "pas pour support l'étendue", which instead we find unchanged in De Mauro' Italian version: "non hanno per supporto l'estensione".

[57] Harris's translation of the French "elles n'ont pas pour support l'étendue; leur siège est dans le cerveau; elles font partie de ce trésor intérieur qui constitue la langue chez chaque individu" seems to imply that language in the mind takes the form of associative relationships. In the Italian translation, De Mauro proposed a weaker interpretation of this passage; although maintaining the focus on the individual, De Mauro does not seem to imply that such associations are located in the brain nor that the form of the language in the mind is reducible to these kind of associations (Saussure, 1967, p. 150)

psychological nature of the *associative* relationships in the "indeterminate order and indefinite number" of the possible associations that leads to a reconsideration of the linguistic notion of *value* (1997, pp. 36–37). Given that *associative* relationships can be of an individual kind, we ought to consider whether these associations are actually a property of the individual or if they can reveal similar paths in speakers of the same language (Basile, 2001, p. 78). On the one side, language is open to the continuous creation of new and unpredictable associations; on the other side, language contains a set of already sedimented relationships (i.e. semantic relationships) to which the speakers continuously refer in their act of communication and in the formation of thoughts (cf. Casadei, 2003a, p. 52). Hence, at least at a very general level of analysis, associative relationships are both individual – in the sense that they depend on idiosyncratic experiences – and systemic – at least in the sense that repeated uses of such associations, and, hence, their standardization within a community of speakers, give rise to stable semantic relationships.

One last aspect we want to outline is that *syntagmatic* and *associative* relationships are not completely independent as they work together in defining the relational *value* of the linguistic unit:

> a linguistic unit may be compared to a single part of a building, e.g. a column. A column is related in a certain way to the architrave it supports. This disposition, involving two units co-present in space, is comparable to a syntagmatic relation. On the other hand, if the column is Doric, it will evoke mental comparison with the other architectural orders (Ionic, Corinthian, etc.), which are not in this instance spatially co-present. This relation is associative (Saussure, 1983, pp. 249–250).

Within the relational framework outlined by Saussure, it seems safe to conclude that the meaning of a linguistic unit depends on its position on the *syntagmatic* as well as *associative* axis:

> La dimensione sintagmatica e quella paradigmatica costituiscono le dimensioni principali attraverso le quali si organizzano e funzionano le lingue, dal momento che ogni parlante quando compie un atto linguistico seleziona degli elementi dal piano paradigmatico e li combina insieme su quello sintagmatico (Basile, 2019, p. 43).

Critical readings of Saussure's theories have stressed the impossibility of reducing meaning to purely differential relationships leaving room for the reintegration of positive elements. The meaning of a word should to be understood as a combination of the relational and positional *value* of the word and of the positive (in the sense of not differential, not relational) aspects that become unavoidable once, in order to understand the kinds of relationships that words entail, we clear the way to the individual and to its mental activities, i.e. to those biological and cognitive mechanisms

that impose positive elements to the system. This interrelation of positive and relational aspects of meaning and the processes of standardization of associative relationships into semantic relationships will remain an open question and a matter of debate throughout this chapter (and even beyond).

In conclusion, the lesson of Saussure has various merits: in the first place, it highlights that, in order to understand and describe the meaning of words, we need to focus on the relationships that words entail with each other and that these relationships are observable within the language system. In the second place, the systematicity of language does not imply the complete denial of the cognitive sphere; quite the opposite, the cognitive dimension becomes fundamental for understanding and studying semantic relations. In the third place, Saussure's semantic ideas led the way to the development of structural semantic approaches that, on the one side, radicalized the Saussurian positions towards a purely autonomous view of semantics and, on the other side, gave birth to some fundamental notions and approaches, such as the notion of semantic field and the compositional approach.

3.2.2. The breakthrough in the study of the structure of language: associative, lexical, and conceptual fields

Structural semantic approaches cannot be reduced to a single paradigm of research and not even the names of the major representatives can easily be summarized here. In general, the common thread underlying this wide family of approaches is their self-declared inspiration to the work of Saussure and the consequent development of some of his main ideas. As we have said before, structural semantic theories tend to radicalize the autonomy of semantics, the anti-referential and anti-psychological nature of meaning, the demarcation between linguistic and non-linguistic knowledge. As it will become clear, not all theories emphasize and underline the same aspects: some authors, as for example theorists behind the notion of semantic field, remained closer to the Saussurian lesson and acknowledged the psychological and social basis of meaning; others, like Hjelmslev and others that followed his traces, tended towards a more logical and formal approach; others, like Lyons, shared the Saussurian theoretical framework but emphasized its limitations and tended, at least to some extent, to re-establish the role of the referent in the semantic analysis (Diodato, 2013, p. 29). Given the complexity and vastness of the history of structural

semantics, we do not aim at presenting a unified and comprehensive account of the various theories.[58] We will rather try to identify and pinpoint the theoretical and methodological innovations that boosted and strengthened the interest for the study of the structure of the lexicon.

The structuralist attention for the organization of the lexicon takes two main paths; on the one side, theorists of the notion of semantic (or lexical)[59] field rest on and expand the Saussurian notions of system and of *associative* and *syntagmatic* relationships. On the other side, Hjemslev foresees and develops the possibility of applying the componential analysis used for phonemes, understood as a set of distinctive features, to the level of the semantic content. We will now briefly turn to each of the above-mentioned lines of inquiry and we will try to condensate some of their major innovations, if considered from the point of view of the aim of our research.

The notion of semantic field finds its first explicit formulation in the work of Gunther Ipsen (1924) - in which we find the famous metaphor of language as a mosaic - and culminates in the neo-Humboldtian approach of Josh Trier (1931) and Leo Weisgerber (1926, 1950, 1954), which resumes the Humboldtian notion of *innere Sprachform* and integrates it with the Saussurian intuitions. As Horst Geckeler notes, the ideas of totality, articulation and structure are the leading forces in the theorization of the notion of semantic field (Geckeler, 1979, p. 80). Language is essentially an "intermediate world" (*Zwischenwelt der Sprache*) that creates a bridge between the individual and the world by means of establishing a language-specific picture of the world (*Weltbild* o *Seinbild*).[60] Within this theoretical horizon, the lexicon is a macro-system articulated

[58] For a comprehensive history of structural semantic approaches see Ščur (1974), Lepschy (1997), and Geeraerts (2010).

[59] In the theorization of the notion of field, Trier left the terminological issue quite undefined by using, in his work, a variety of different and often overlapping terms; *Wortfeld* indicated the lexical field, *Sinnfeld* and *Begriffsfeld* the conceptual field, *Sinnbezirk* the semantic sphere but it remained overall undefined. Nowadays, the notions of lexical and semantic field are generally used interchangeably to indicate the set of lexemes that covers a certain conceptual sphere although with sensible oscillations from one approach to the other. Such terminological uncertainty is not just a matter of scientific agreement, but instead it highlights a deeper theoretical issue (especially when it comes to the distinction between lexical/semantic field and conceptual field) concerned with the difficulty in distinguishing between the linguistic and the conceptual levels (cf. Casadei, 2003a, p. 63).

[60] In the Neo-Humboldtian perspective of Trier and Weisgerber, each language selects and conceptualizes what is pertinent for a certain community at a certain historical time. Hence, each language constitutes a different reality and provides the speaker of a certain community with a complete picture of reality. Thus, semantic structures (as well as grammar and syntax) reveal the way in which speakers give form to their reality. Humboldt and Saussure are the direct precursors and theoretical source of Trier's and Weisgerber's argument, who further developed the idea of the formativity of language. However, the main heritage and innovation of these scholars has to do with the notion of semantic field and thus with the idea that the lexicon has to be understood as a web on interconnected relationships within the language.

in micro-systems (the *lexical fields*), which are linguistic entities collocated between the word and the whole of the lexicon (Diodato, 2013, p. 30).

Trier starts from the structuralist insight that only a mutual demarcation of the words under consideration can provide a decisive answer regarding their exact value. Hence, the first merit of the Trier-Weisgerber notion of field is to have shown that the lexicon cannot simply be a list or a chaotic mass of non-organized words, but that it is rather made up of structured and interconnected subsets. The meaning of words cannot be grasped if we consider words as single, atomic, and unrelated entities, but, instead, we ought to look at the structure in which words are integrated and at the web of semantic relationships they entail. The substance of human knowledge—the contents of cognition—is divided by language into a number of adjoining small areas, in the same way in which a mosaic divides a two-dimensional space by means of contiguous mosaic stones. According to Trier, words are grouped together in fields in which each word is semantically related to the others by means of *associative* (*paradigmatic)* relationships. In this sense, Trier hypothesizes that the relation by means of which language is structured in semantic fields is fundamentally the *paradigmatic* one. In Trier's notion of field, *associative* relationships become standardized and codified as semantic relationships in the *langue*. Hence, semantic fields are not subjective and virtually unlimited psychological constructs based on the free associations that speakers might make between words; rather, they have to be understood as linguistic structures shared by the linguistic community.

Consequently, the notion of semantic field tends towards an autonomous view of language: it is in the semantic relationships (more than in the mere associations between words) that we can grasp the structural aspects of language and, consequently, the value of the linguistic units. Among the virtual and individual associations, there are some that depend entirely on the structure of language "in quanto riconosciute dal parlante come inscritte nel patrimonio collettivo della *langue"* (Diodato, 2013, p. 31).

The notion of *associative field* put forward by Charles Bally (1932) more closely addresses the notion of *associative* relationship as a cognitive phenomenon; in fact, the *associative field* of Bally embraces the constant flow of assimilation, association, comparison and opposition of the elements of language carried on by our minds and, consequently, it is by definition open, undetermined and virtually unlimited. Among the multiplicity and variety of the possible associations, some have to be ascribed to the individual and others seem to have a sort of privileged

and stronger role assigned by the fact that they seem to be shared by the majority of the speakers.[61] Hence, *associative* relationships are partly left to the free innovation of the individual, and partly bounded to and conditioned by the system of the language.[62]

We are here highlighting one specific aspect of the notion of field, i.e. how field theory deals with the intricate relationship between *langue* and *parole*. One of the fundamental questions underlying both Trier's and Bally's approaches is whether associations belong to the realm of the *langue* or to the realm of the *parole*, i.e. if they are an effect of individual uses that constitutively involve other types of knowledge, or if fields have to be thought of as based on purely linguistic-semantic relationships and, consequently, of being a specifically linguistic fact. We have briefly mentioned the view of the two authors. If we assume semantic fields to be autonomous linguistic facts, then the question is how the linguistic level of organization relates to the non-linguistic level of our knowledge. To this regard, one of the most often mentioned pitfalls of Trier's notion of semantic field is exactly the nature of its relationship with the conceptual field. Although central to us, this sort of question was not (and could not be) of central importance for Trier, inasmuch as one key argument at the very outset of structural semantics was to develop a non-cognitive (or even anti-cognitive) account of meaning. Nonetheless, we can draw from some parallel observations to understand how semantic fields may relate to non-linguistic cognition.

Allow us one further preliminary observation: Trier developed the notion of semantic field to explore and understand the diachronic change of the meaning of words. According to Trier, one of the main issues with the diachronic study of meaning was the analysis of isolated words, without taking into consideration how changes occurred within the structure of the lexicon. Trier's suggestion was that, in order to understand how meaning changes, we ought to confront how distinct semantic fields – semantic fields that are distinct over time - cover the same conceptual

[61] As Basile clearly highlights, Bally envisages the possibility of some associations to be so preferential and privileged to generate an internal and necessary association. Such associations would be at least partly motivated within the language system. According to Bally's view, the strength of the association depends on the motivation of the sign within the language system (Basile, 2001, p. 79).

[62] According to Geckeler, Bally's approach does not meet the requirements of a structuralist semantic approach; the associations that Bally wants to include in the *associative field* represent various kinds of semantic relationships that in part go beyond the realm of language and are grounded in the non-linguistic context. Bally does not operate a clear distinction between linguistic facts, i.e. the words and their interdependent relatonships, and facts of the extra-linguistic reality that are not organized by means of oppositional relationships. Hence, the foundation of a structural semantics needs to leave out associative relationships, that although remain useful in the analysis of the texts (Geckeler, 1979, pp. 124–125).

space[63]. It is not the meaning of single words that is modified over time. On the contrary, the organization of the semantic field is progressively restructured every time we introduce a new term or modify an existing term, and the restructuring of the lexical field leads to a reorganization of the conceptual space that corresponds to the modification of the social and cultural environment.

According to Violi, one pitfall in Trier's approach is the constant and ambiguous swing between the analysis of the semantic field (considered as a linguistic fact) and of the conceptual field to which the organization of the various sematic fields is compared. As Lyons argues, the distinction between lexical and conceptual field cannot be directly ascribed to Trier's position (Lyons, 1977, p. 253). In fact, as we have already mentioned, coming from a Humboldtian and Saussurian background, Trier claims that it is the "organic and interdependent articulateness" of language that shapes and determines the components of the conceptual field. Nonetheless, even though Trier aims at comparing linguistic expressions within the framework of a structuralist approach, he is often forced to refer to a not-well-defined conceptual space[64] in order to describe differences in the semantic field (Casadei, 2003a, p. 64; Violi, 1997, p. 42). Consequently, the theoretical problem consists in understanding what is meant by saying that something is or is not the same conceptual field. Within the framework of Trier's approach to the notion of field, we find no straightforward (nor latent) definition of this identity; and yet this is the constant, in relation to which changes in the meaning of lexemes in diachronically distinct lexical fields are determined (Lyons, 1977, p. 258).

To this regard, it is worth mentioning the debate that occurred among supporter and critics of the notion of field about whether this notion is more suitable to the study of concrete rather than abstract domains. Trier endorsed the idea that the notion of field would find its most proper

[63] It is important to point out that the diachronic aspect of language change was the focus of Trier's research and the issue that led him to the theorization of the lexical field. Trier studied the changes occurred in medieval German in the domain of knowledge and intellectual activity; at the beginning of the 1200, such field was organized into three main words, *wisheit, kunst* and *list*, and the same field in the next century was covered by the terms *wisheit*, *kunst* and *wissen*. By mean of the analysis that the words entailed in the 1200 and in the 1300, Trier showed that the change from *list* to *wissen* was more than a mere substitution, but rather a reorganization of the senses and relationships between the senses connected to the profound cultural changes that occurred with the breakdown of the feudal society (cf. for example Geckeler, 1979, pp. 60–61 for the importance of the diachronic approach and the innovations to which Trier's approach led).

[64] Following the lesson of Saussure, Trier seems to endorse the idea of an amorphous matter of concepts previous to its organization in language (Diodato, 2013, pp. 29–35). The lexical field theory is based on the structuralist assumption that meaning is a linguistic entity and that different languages impose their peculiar organization to an undifferentiated prelinguistic world and, hence, that concepts are essentially linguistically defined. Taken as a strong claim about the nature of the relationships between language and thought, this perspective can be linked with a strong deterministic claim often associated with the principle of linguistic relativity (cf. Casadei, 2003a, p. 63).

application in the analysis of abstract fields; on the contrary, most lines of research that developed the field approach applied the notion to concrete fields, where the lexemes have identifiable denotata. The analyses of concrete fields (might) allow to by-pass the issue of the identification of the conceptual space for contrasting among different fields. On the one side, it seems fair to argue that changes in the lexical field, and hence in the conceptual field, are easier to track if we can constantly check the organization of the field against a quasi-objectively described external reality – we can think again about the domain of colors. On the other side if we formulate the notion of field within a non-conceptualist framework, "[t]he truth of the matter seems to be that the determining principles of lexical structure apply equally to both abstract and concrete words" (Lyons, 1977, p. 281).

A similar circularity between lexical and conceptual fields seems to arise when we try to define the boundaries of a semantic field or when we assume that words completely cover a conceptual space without leaving gaps. Two important criticisms to Trier's field theory became the starting point for important developments and innovations in the semantic field approach. They can be summarized as primarily concerned with the internal constitution of the lexical field and with its external boundaries.

On the one side, Trier admitted the difficulties in establishing the external boundaries of a semantic field: it is difficult (if not impossible) to identify a precise criterion that would determine where a certain semantic field ends and where another begins. However, this does not pose intrinsic limitations to the study of fields. The articulation of the semantic field can be studied even without the characterization of the actual external boundaries (cf. Geckeler, 1979, p. 110) and the areas in which the lexicon is internally organized are not defined and fixed, but rather flexible and with vague boundaries (cf. Violi, 1997, p. 43). On the other side, Trier seems to adopt the metaphor of language as a mosaic (introduced by Ipsen) to explain the relationship between the conceptual and the lexical field; according to this metaphor, words entirely cover a given conceptual space without leaving gaps[65] and hence, if we interpret the notion according to its stronger version, the whole of the lexicon is composed of adjacent and consecutive fields. Such definition underestimates how pervasive lexical gaps might be and does not provide a clear understanding or a convincing account

[65] *Lexical gaps* occur when the coverage of the conceptual field by the lexical field is not complete: while *horse* is a cover term for *stallion* and *mare*, no similar term exists for *bull* and *cow* (Geeraerts, 2010, p. 56; Lyons, 1977). A *lexical gap* can be defined as a missing item only in relation to a lexical field. For various possible kinds of *lexical gap* see Aitchison (1987) and Diodato (2013).

of how language delimits and organizes the conceptual sphere. Trier himself does not seem to fully embrace the metaphor of the mosaic and leaves room to the possibility that language only apparently covers the whole of reality (Diodato, 2013, pp. 32–33).

Going back to our original question – if semantic fields ought to be taken as linguistic facts or if they involve our types of knowledge, i.e. if semantic fields are a matter of *langue* or of *parole* – the criticisms that we have reviewed above seem to be implicitly based on the idea that *langue* and *parole* are two distinct aspects of language and on the assumption that one can be studied without considering the other. Hence, returning to the nature of associative relationships and following our argument in the previous chapter, we can see how it is only in the dialectic relationship between *langue* and *parole* that we can understand and study meaning and its organization. It is only by means of the continuous use of language that we progressively standardize repeated associations into stable relationships. The double and interconnected nature of the mental association (i.e. being both a matter of individual uses and of standardized practices) allows to understand the nature of semantic relationships. Semantic relationships are the stable and linguistically codified side of the associations:

> le associazioni mentali cederebbero, per così dire, il passo alle vere e proprie relazioni semantiche le quali riguarderebbero non esclusivamente il piano dei concetti ma, in un orizzonte più specificatamente linguistico, il piano dei significati intesi saussurianamente come classi astratte di significazioni, come un *ubi consistam,* come "zone" di stabilità a cui ricondurre tutto ciò che è più fluido e variabile (Basile, 2001, p. 79).

The semantic or lexical field is an important principle of organization of the lexicon, according to which words that belong to the same activity or portion of knowledge tend to be grouped together in the mind of the speakers in an integration of linguistic and non-linguistic knowledge. This means that, despite their theoretical limitations and the structuralist background in which they were elaborated, the fundamental idea that we aim at bringing forward in our argument is that semantic fields are a particular way in which we structure a certain conceptual space by means of an open set of lexemes. This is often viewed as the main revolutionary aspect that comes with the introduction of field theory and the reason why field semantics is still a widespread approach (even if in highly revisited forms) that found interesting applications and validations in areas or research that distance themselves from structuralist assumptions, such as psychology and cognitive

sciences. The notion of semantic field is fundamental to analyze the lexicon as a combination of interconnected subsets. What constitutes the cutting edge of the semantic field approaches is the possibility of studying portions of the lexicon by focusing on the relationships entailed between words (Diodato, 2013; Lehrer & Kittay, 1992).

3.2.3. Meaning is not just by associations: semantic syntagmatic fields

We will conclude this brief exploration in the notion of field in linguistics by mentioning another set of fields that received overall less attention in the development of structural semantic accounts: *semantic syntagmatic fields*. We have seen how both Bally's and Trier-Weisgerber's theories of field was mainly concerned with *associative* (*paradigmatic*) association. As for syntagmatic relationships, it gradually became clear that words may have specific combinatorial features and that such features had to be included in the analysis of the field. Traditionally, the possibility of combining words with other words was looked at mostly from a grammatical perspective; for example, if a word belongs to the grammatic category of the verb, then it implies a noun as its subject.

Walter Porzig (1934) introduced the notion of *semantic syntagmatic field*[66] and defended the importance of the notion of *elementary semantic field (elementare Bedeutungsfeld)* to explain the semantic relationships entailed between combinatorial bipartite structures composed by, for example, a name and an adjective or a name and a verb, such as *to see – eye, dog – bark, neigh – horse* (cf. Geckeler, 1979, p. 76). Porzig pointed out that "syntagmatic combinability has as much to do with aspects of meaning as with grammatical characteristics" (Geeraerts, 2010, p. 58). It is an "essential" and "purely linguistic" semantic relationships that links the words of the pairs mentioned above, given that it is almost impossible to explain the meaning of a lexeme without involving the meaning of the other (Diodato, 2013, p. 34). In Porzig's notion of syntagmatic field, we can recognize a first case of *collocation*: some lexemes occur in combination with other lexemes since, on the basis of a relationships of semantic implication, starting from the presence of one we can predict the presence of the other. Other cases of collocation include lexemes that,

[66] Jolles (1934) proposes a theory of field similar to that of Porzig. Jolles notion of *Bedeutungsfelder* is narrowed to the analysis of fields composed by two lexemes (e.g. *father – son, day – night*) and, therefore, it was limited to immediate oppositive pairs, whereas Porzig definition of field dealt with combinatory structures. For this reason, Jolles theory did not have successive developments in structural semantic approaches.

although they are not based on an "essential" semantic relationship, tend to occur in fixed syntagms even though they are potentially free to occur in a plurality of different combinations. We defer the specific analysis of Porzig's theory to the relative section in Lyons's *Semantic* (Lyons, 1977, pp. 261–266) and we will only report some general considerations. The theoretical significance of Porzig's argument lies in the recognition that syntagmatic relations, no less than paradigmatic relations, determine the structure of a lexical field and are fundamental in order to understand the meaning of words. For example, there are many distinctions of meaning that can be made either by the syntagmatic modification of a more general lexeme or by the use of a more specific single lexeme. Hence, associative and syntagmatic relationships are both fundamental in the analysis of the structure of the lexicon: both reveal organizational aspects of the lexicon and, therefore, have to be integrated in a lexical-semantic theory.

The complementarity of paradigmatic and syntagmatic approaches was central in the development of Eugenio Coseriu's theory[67], who, starting from Porzig's proposal, developed the notion of *lexical solidarity*. In his works, Coseriu developed a comprehensive classification of various lexical structures and characterized the lexical field within a more general account of the structure of language (for a review of Coseriu's theory see Geckeler, 1979, pp. 141–164). In Coseriu's view, the system of language is organized in lexical (or lexematic) structures. The lexical structures can be divided into *paradigmatic* and *syntagmatic* ones. The first are based on relationships of opposition at the *paradigmatic* level, i.e. *paradigmatic structures* are constituted by the lexical units that might fit into the same space of the written or spoken chain and that, therefore, constitute a paradigm, a system of oppositions. *Syntagmatic structures* are based on the lexical solidarities and are, therefore, combinatorial structures. The notion of *lexical solidarity* is defined as a linguistically codified syntagmatic implication at the level of the content such that one of the terms (as for example *nose*) is a distinctive feature of the second (*hooked*), i.e. the first is implicated in the definition of the latter. Coseriu notes that the *lexical solidarity* is generally directed: *nose* is included in *hooked* from the point of view of the content, but not the opposite (Coseriu, 1967, trad. ita. 1971a; Geckeler, 1979; Ježek, 2005, p. 170).

[67] We will come back to Coseriu's approach and integration of the lexical field and compositional approaches in section 3.5. Our aim at this point of the argument is to focalize briefly on the necessity of integrating paradigmatic and syntagmatic observation in order to develop a comprehensive account on the meaning of lexemes within (as well as outside) a structuralist perspective.

Coseriu polemizes with Porzig's position for not having clearly distinguished between syntagmatic implications (i.e. *lexical solidarities*) and those implications provided by an encyclopedic knowledge about the world. Implications like *the snow is white* or *muddy*, or *the dust is gray* or *dry* - i.e. the fact that a noun may have properties such as color, form, and dimension - has little to do with the relationships entailed between the content, such as the implication between *dog* and *to bark*. Furthermore, Porzig did not distinguish the various types of solidarities that can be observed between words. According to Coseriu, it seems highly problematic that all "essential semantic relation" belong to the same lexical phenomenon; conversely, he came to propose a distinction between three possible mechanisms that underlie syntagmatic (combinatory) structures; affinity, selection and implication. But more than the distinctions themselves, it is interesting to note the theoretical ground that allowed Coseriu to come to such distinctions. The analysis of the combinatorial structures is based on notions that are, at the same time, implied in the definition and description of the paradigmatic structures (especially of the *lexical field* and of the *lexical class*), such as the notions of *lexical field*, *lexeme*, *arcilexeme*, *class*, and *classeme*. This approach, that we will not further investigate in its more technical aspects, substantially provides an integrated perspective on the structure of language, that encompasses both paradigmatic and syntagmatic relationships within a unified theoretical framework.

To us, this last consideration is of fundamental importance for the development of a full-blown account of the structure of meaning: the structure of the cannot be grasped by only looking at the associative (paradigmatic) relationships between words, which constitute only one aspect of the meaning of words. In their original formulation, the lexical field is a kind of paradigmatic structure and does not cover the entire structure of the meaning of words: it is one of the articulations of the structure of the lexicon. We cannot avoid pointing out the general terminological confusion in the way in which we, including our discussion, talk about the relationships that can be established between words, especially when we try to expand the notion of field to syntagmatic relationships. Talking about semantic or lexical organization from a mere paradigmatic perspective seems indeed a reductive move. The study of the meaning of words and of their organization within language has to encompass the limitations and determinations that come from the syntagmatic level of the organization of language. We will come back to this terminological issue towards the end of our discussion. By means of this observation and of what we previously said about the nature of associative and syntagmatic relationships within the

framework of the Saussurian view of meaning, we merely scratched the surface of this fundamental integration of syntagmatic and paradigmatic aspects of meaning, that will become one of the focal points of interest when we turn to the distributional accounts of meaning (chapter 5).

3.3. Structuralist approaches to the structure of language: relationships "within" words

3.3.1. The rise of compositional approaches to meaning

In the previous sections, we have mostly been concerned with the problem of identifying and describing the structure of meaning at the level of the relationships entailed between multiple words. But this is not the only way to look at the structure of language. Linguistic units themselves seem to exhibit a structure identifiable below the level of the word. In this section, we will not specifically engage with debates over the structure of the lexicon, but rather we will turn to those approaches that discuss the structure of linguistic categories and, hence, the structure of meaning.

Within structuralist approaches to meaning, one important theoretical as well as methodological novelty is the componential analysis on meaning, according to which the meaning of a lexeme can be analyzed by means of a set of basic components (or features). Even though this theory was originally formulated independently from the notion of field and, as we said, does not directly address the structure of the lexicon, but rather the structure of the linguistic category, it is in the combination of the two methods that we see some of the most fecund approaches. However, already in its independent formulation, the componential analysis of meaning has been and (to some extent still is) at the base of a variety of approaches both in linguistics[68] and in the cognitive sciences[69].

In the framework of our research, the discussion of compositional semantic approaches is interesting because it allows us to discuss topics related to the traditional notion of category (and specifically linguistic category), according to which categories have clear-cut boundaries and can be described by a set of properties that define them and that entities have to share in order to be

[68] For some of the most relevant representative of the compositional approach in the European research environment are for example Greimas (1966), Prieto (1964), Coseriu (1958, 1964, 1971b, 1973).
[69] In the development of our argument we will touch on some of the applications and reformulations of the original compositional approach, with a specific attention to those lines of inquiry arisen from and within the cognitive sciences. For a comprehensive discussion and review of the literature see (Geeraerts, 2010).

part of such or such category (Diodato, 2015). Given the amount of literature and research that is linked to the idea of the compositionality of meaning, we will not attempt to retrace its history or to give a comprehensive overlook on the various approaches. On the contrary, we will try to outline some of the main points of interest risen by these approaches. After a brief introduction to the basic assumptions underlying the compositional approach, we will mainly focus on the kind of structure of language it envisions and on the fundamental debate about the nature of the semantic features. As it will become clear through our discussion in this and later chapters, a traditional notion of linguistic category does not seem to explain much of the most recent research in language and cognition. We will try to sketch how researches progressively moved away from this traditional notion and how more recent views on the nature of conceptual and linguistic categories may help us shed light on the organization of language.

One of the leading figures of the compositional approach, in its derivation from the Saussurian principles,[70] is Louis Hjelmslev. As we have seen in the previous chapter, the Saussurian notions of language as a system and of language as a form receive particular attention in the development of Hjelmslev's theory. Under the influence of theoretical approaches in formal logic, Hjelmslev retrieves the Saussurian notion of language by reformulating and integrating Saussure's principles within a rigorous and abstract description of the language system. Within this theoretical framework, the study of meaning and, more precisely, of the semantic structure of the linguistic units, becomes a fundamental aspect of Hjelmslev's proposal.

Starting from the Saussurian principle of the bifaciality of the sign, Hjelmslev's hypothesis is that the two levels of the sign are structured on the base of the same principles - principle of the isomorphism of the expression and of the content or of the symmetry of the sign.[71] On the level of the expression, each language selects some classes of sounds (phonemes) and not others. Similarly, on the level of the content, each language gives rise to a particular set of conceptual distinctions. The next move was to apply the methodological innovations of the School of Prague on the matter of phonology to the level of the content. As Geckeler points out, Hjelmslev transposed to the analysis of the lexicon the idea that in order to study the phonology of language we need to go

[70] Even though Hjelmslev's theory is directly inspired by the work of Saussure, it often brings the Saussurian principles to its extremes (as for example when it comes to the idea of the autonomy of language), which may not have been contemplated by Saussure in such ways. For example, it does not seem to see directly in Saussure the precursor of a compositional account of meaning (Diodato, 2015, p. 12 note 6).

[71] We will not discuss here the theoretical adequacy of such hypothesis for the discussion of which we refer to Albano Leoni (2009) and De Palo (2003).

under the threshold of the sign (Geckeler, 1979, p. 166). This general idea is actualized in the application of the definition of phoneme as set of distinctive features to lexemes. Hjelmslev recognized that the unlimited number of the linguistic signs stems from the combination of a limited number of "non-signs" called *figurae*:

> a language is so ordered that with the help of a handful of figurae and through-ever new arrangements of them a legion of signs can be constructed. If a language were not so ordered it would be a tool unusable for its purpose. We thus have every reason to suppose that in this feature – the construction of the sign from a restricted number of figurae – we have found an essential basic feature in the structure of language (Hjelmslev, 1961, p. 47).

Some of the most relevant notions imported by Hjelmslev from phonology to semantics are the principle of reduction, the notion of semantic features, and the commutation test. The principle of reduction is essentially summarized in the above citation: the main idea is that it is possible (and necessary) to decompose the atomic parts of language (lexemes) into sub-atomic and elementary particles, both on the plane of the expression as well as on the plane of the content (cf. Geckeler, 1979, p. 166). Thus, the aim of semantic analyses is to determine this limited set of elementary, non-further-decomposable particles. These sub-atomic particles are the *figurae* (in Hjelmslev's terminology) or what we would today call, for the plane of the content, *semantic distinctive features*. Such features are invariants if "the exchange of one and only one element for another is in both cases [both of the plane of the content and on the plane of the expression] sufficient to entail an exchange in the other plane of the language" (Hjelmslev, 1961, p. 70). If, in the product of +HUMAN +ADULT +MALE, we exchange +MALE with -MALE we obtain a change from *man* to *woman*. This method is called commutation test, which was originally formulated by Trubeckoj in the study of phonology and was then extended by Hjelmslev to the plane of the content. This kind of analysis has to be conducted by means of the continuous correlation between the plane of the expression and that of the content; the test shows how and if the substitution of an element with another on the plane of the expression leads to modifications on the plane of the content and vice versa.

As we can easily see, these principles seem to coherently develop the structuralist assumption of the autonomy of semantics and of the intra-linguistic nature of meaning (Violi, 1997, p. 85). This concise reconstruction of the main theoretical principles of the compositional

approach allows us to highlight some of its most problematic points (at least from the theoretical perspective we have been setting out so far). The possibility of decomposing the meaning of a lexeme into a set of elementary component implies that we need to identify and characterize such components; in a sense, compositional accounts of meaning move the problem of the nature of meaning from the atomic level of the word to the sub-atomic level of its components. Hence, we then need to ask what the nature of these semantic components is and how they cooperate in determining the meaning of words.

One central aspect of these invariants is their dichotomic nature: semantic distinctions can be represented by means of dichotomic and privative terms, and as a function of the presence or absence of a certain semantic feature (for example, in *man* and *woman* the distinction between the terms arises from the presence or absence of the feature +MALE). The principle of the binary nature of the semantic features poses some fundamental problems: on the one side, as Lyons clearly shows, we could ask why a *woman* or a *girl* lack of the +MALE component and not the opposite (i.e. that a *man* or a *boy* lack of the +FEMALE component). In the process of decomposition of many antonyms or complementaries this would not be a problem "if it were not for the fact that in other pairs of complementaries, like 'cow':'bull' and 'duck':'drake', it is the one denoting the male that is semantically marked" (Lyons, 1977, p. 322). Thus, the general problem is that of determining which term in a pair of complementaries is the marked one and which is the unmarked. Even more problematic is the application of the notion of binarity of the features to those lexical contrasts that are not dichotomic at all or to properly identify a notation for those words that are unmarked in regards to a privative opposition (such as the English word *horse*, with is compatible with both +MALE and -MALE) and words like *house* that is not compatible with neither (cf. Lyons, 1977, pp. 352–353).

The general underling issue that we can glimpse at by means of these few examples is the problem of determining which components have to be included in the list of attributes that composes the meaning of a word and, hence, what kind of meaning we describe by means of a compositional approach. In a sense, we need to face the same problem we have encountered before: the intrinsic difficulty of separating linguistic from the non-linguistic knowledge when analyzing the meaning of words.[72] Hjelmslev's hypothesis sets forth the theoretical possibility of identifying

[72] For an in-depth discussion of the problematic separation of linguistic and non-linguistic knowledge in compositional semantic approaches cf. Eco (1984) and Violi (1997).

purely linguistic distinctive features, putting aside all non-linguistic aspects. Attempts made in the compositional literature to maintain the level of language separated from the level of experience have all ended up showing the line of continuity that links these two, instead of the line that should separate them (cf. Diodato, 2015, pp. 15–16). As Violi states,

> È evidente, infatti, che ciò che noi sappiamo di un termine come *uomo* è estremamente più complesso dell'intersezione di MASCHIO, UMANO, ADULTO, ANIMATO; virtualmente comprende tutte le conoscenze che possediamo dall'uomo di Neanderthal fino alla costituzione fisico-biologica dell'organismo umano. [...] Naturalmente la linea argomentativa scelta dai sostenitori delle semantiche a tratti è che non tutte queste conoscenze sono *costitutive del significato* [...]. Il problema, come è ovvio, è *come* stabilire cioè che è costitutivo e ciò che non lo è (Violi, 1997, pp. 86–87).

The application of the compositional semantic approach to historical-natural languages[73] has shown that, on the one side, the number of features increases in proportion to the number of the lexemes that are analyzed and, on the other side, that it is often necessary to include features that are not limited to the dictionary meaning of the word, but that rather have to specify the grammatical category, its restrictions in the collocations, or its pragmatic aspects (cf. Diodato, 2015, p. 16).

The second main problematic aspect of the compositional view of meaning has to do with the nature of the semantic features. According to Hjelmslev, these semantic features do not have an ontological status and should not be understood as universal atomic concepts, but rather as components of meaning distinct from the lexical units. The main idea is that these semantic features compose a sort of metalanguage useful to the description of meaning, and that, at the same time, they do not require to be taken as psychologically or ontologically real. It should be noted that assuming the psychological or ontological reality of semantic features would be in direct contract to the structuralist assumptions of the autonomy and anti-psychological nature of meaning (cf. Diodato, 2013, p. 28). Hence, semantic features are not part of our mental components, and do not have a mental counterpart. At the same time, they are not lexical units. This last point is actually quite controversial and has been pointed out by De Mauro; in fact,

[73] In the next sections we will discuss researches that took on such endeavor both in linguistics (Alinei, 1974) as well as at the intersection of linguistics and psychology (Wierzbicka, 1985, 1996).

> i tratti distintivi dei fonemi non sono essi stessi fonemi: sono entità d'altro ordine, di numero finito, che definiscono i fonemi e sono definiti dalla fonetica e dall'acustica; invece i presunti tratti distintivi dei lessemi sono essi stessi parole (umano, bovini, adulto, ecc.) e, come tali, abbisognano essi stessi di essere definiti da tratti pertinenti che sono di nuovo parole, in una fuga senza fine che aggiunge di continuo tratti alla lista di tratti rendendola irrimediabilmente una lista aperta, tale quale la serie di lessemi di una lingua (De Mauro, 2008, p. 67).

As we will see, the problematic nature of the semantic features will become a central question in compositional accounts of meaning stemming from a cognitive background.

A third problematic aspect of the theory is the underlying notion of linguistic category on which compositional accounts rest. If we consider the semantic traits as necessary and sufficient conditions for the definition of the linguistic category, no features can be added or deleted. At the same time, all features have the same definitory value and the same relevance in the definition of the category. Furthermore, within a linguistic framework, semantic features give an exhaustive account of meaning that is the resultant of the product of the distinctive features; hence, the definition of meaning is not gradual but rather determined by clear-cut boundaries (cf. Violi, 1997, pp. 84–85). This definition of linguistic category rises issues both when considered from a cognitive perspective, as we will later discuss when evaluating the cognitive plausibility of such notion, as well as within a linguistic perspective as the one outlined in the previous chapter. In fact, if meaning is understood within the framework of an experience-semantic-pragmatic *continuum,* and as a web of pertinences that organizes the multiplicity of the sense and acceptations in which words are used, then a traditional definition of category, and especially the idea that these categories have clear-cut boundaries, could hardly be incorporated and accepted as a plausible theory of the structure of meaning.

However, this does not imply that a compositional account of meaning has to be ruled out completely. Heavier criticisms to the plausibility of compositional principles will arise within the cognitive semantic approach and will especially be concerned with the compositional account of meaning developed within the generativistic framework. This discussion of the problematic aspects of compositional approaches give us the stepping stone to introduce the next steps of our argument; with the rise of cognitive linguistics, one of the fundamental questions in the exploration of the structure of the lexicon became the issue of the cognitive plausibility of such theories. This problem is not inherent to the theories we have explored so far, which, as a matter of fact, had as

a founding principle the fact that language could and should be studied within itself. Nonetheless, the problem of the cognitive plausibility becomes a matter of reflection both as we move on with our historical review as well as within the framework we outlined in the previous chapter and in which we are setting our discussion. But, first, let us look at what happens if we combine field theory and compositional accounts of meaning within a linguistic framework.

3.3.2. Integrating lexical field and compositional approach

Before turning to the more specific cognitive aspects of the debate delineated so far, we will very briefly introduce the reader to some of the most relevant approaches that integrated the notion of lexical field and the compositional approach within a structuralist perspective. Although the two approaches were born off of some of the same Saussurian assumptions but, at the same time, developed two different lines of research, it did not take long to recognize that they were the logical integration one of the other. The compositional analysis becomes the theoretical framework to reduce the internal as well as external blurriness of the original notion of semantic field. In very general terms, according to supporters of the componential approach, words that share one or more distinctive features are semantically related, and words that share at least one semantic feature form a lexical field. The content of a lexeme results from the structure of its semantic features and the configuration of the semantic field becomes a proper linguistic structure only by means of the existence of distinctive oppositions; the main idea is that the lexical elements in the field are distinguished by functional oppositions. Hence, it can be argued that all words that share at least one distinctive semantic trait form a lexical field.

Overall, this short introduction does not aim at theoretically contributing to our discussion, but rather at briefly reviewing some relevant applied works with the hope that they will clarify some of the theoretical standpoints mentioned in the previous sections. Hence, we will consider two main approaches that, as we will see, are concerned with slightly different aspects of the compositional and field theory interaction: Coseriu's *Lessematics* and Alinei's investigation of the semantic domain of the *horse* in Italian.

Coseriu further developed the notion of field in a structuralist sense and integrateed it with the compositional accounts of meaning, giving rise to a new paradigm of research called *Lessematics*. Coserius's formulation of the lexical field theory can be seen as a deliberate

methodological attempt to draw the consequences of a structuralist approach to meaning, from the distinction of the proper object of investigation to the framework for the description of the lexical items. One of the main objections to the original definition of lexical field is its lack of a proper linguistic methodology, beyond the intuition of the researcher (Geckeler, 1979, p. 141). The principles underlying the foundation of Coseriu's methodology are: the necessity of keeping linguistic reality distinct from non-linguistic reality and the focus on the functional organization of the system (the *langue)*, which comprises all that is distinctive in language.[74] According to Coseriu, and in line with the structuralist tradition, language and the world are not structured in the same way. Words do not just reproduce the distinctions that we can observed in the world, but they trace limits and boundaries that are proper to the language system. As a consequence, Coseriu focused on the autonomy of the language system and clarified how the study of lexical semantics has to be strictly concerned with the linguistic values, i.e. with the distinctive oppositions within the relationships of signification (cf. Diodato, 2013, p. 36). This becomes important to recognize that further methodological restrictions apply to the study of lexical fields:

> the definition of 'lexical field' is restrictive to the extent that associative fields are explicitly ruled out. Coseriu admits only fields that consist of lexical items that exhibit clear oppositions, like young and old, day and night, or tiède 'lukewarm', chaud 'warm', and brûlant 'hot', where the items (unidirectionally or bidirectionally) exclude each other (Geeraerts, 2010, p. 79).

Moreover, Coseriu's stepwise observations and restrictions of the object of investigation leads to the point of recognizing that the structural method of investigation does not apply to the whole of the phenomenon of language or to the entire lexicon; terminologies, specialized vocabularies and enumerative structures, that, according to Coseriu constitute mere "nomenclatures" (i.e. they merely name things in reality), do not exhibit the type of mutual distinctive oppositions that structural semantics is interested in (cf. Geeraerts, 2010, p. 78).

Within this framework, the notion of semantic field is integrated with the componential analysis of meaning. According to Coseriu, in order to reach a proper structural method, the lexical

[74] To this regard, we should note the important distinction made by Coseriu between *norm* and *system*. According to Coseriu, the *norm* comprises everything that is not necessarily functional (i.e. distinctive) but that, nonetheless, is fixated by means of the current and common use in the linguistic community of speakers. On the contrary, the *system* comprises what is objectively functional. Structural semantics is mainly concerned with the level of the *system* as the place of the functional oppositions (cf. Geckeler, 1979, pp. 150–151).

field has to be studied in correlation to the compositional method, i.e. the lexical structure hypothesized by Trier and Weisgerber has to be applied to a level below that of the word.

One fundamental innovation of Coseriu's approach is the notion of *arcilexeme*, which solves the problem of the structure of the semantic field. Coseriu recognizes that it is important to distinguish between the positional value of the words within the field and the global meaning of the field; for each field we can identify a *content* or *unitary value* on the basis of which the field is in opposition to the other fields. Such *unitary value* can be actualized in a lexical unit or not, and, thus, provides a methodological definition of the external borders of the field as well as its internal structure. To provide an example: in French we cannot find an *arcilexeme* for the field *vieux, ancien, âgé, jeune, neuf,* but, on the contrary, the *arcilexeme* of the field of the chair analyzed by Pottier (*chaise, fauteuil, tabouret, canapé, pouf)* would be *siège* (Geckeler, 1979, p. 154).

As for the internal structure of the field, the unitary value of the field is a primary distinctive feature on the basis of which we can identify the lower-level distinctive oppositions between the words that belong to a certain field; as for the external boundaries, we can think of a lexical field as the combination of words that share the same dominant distinctive feature (the *arcilexeme*) and each time we encounter a change in the *arcilexeme,* we are confronted with a different field. Following on from the work of Coseriu, Geckeler (1971b, 1971a, 1979) proposed to employ the principle of the field elaborated on structuralist assumptions and integrated with the composition method to propose an adequate analysis of the problem of the synonyms and antonymous, given that synonymy and antonymy can be viewed as content's relationships within the lexical field. Geckeler puts this new method to test by applying it to a limited but complex field, that of adjectives denoting age and allied phenomena in Modern French (Geckeler, 1971b) and later compares the structure of the field *alt - jung- neu* in a contrastive analysis of Modern Italian, French and Spanish (Geckeler, 1971a).

In line with the same intuitions of Coseriu about the fruitfulness of a compositional-field approach, the Italian linguistic Mario Alinei (1974) developed a comprehensive study of the lexical field of the *horse* in Italian. The theoretical aim of Alinei's proposal was to outline the possibility of integrating the compositional principles of the structural units of language and their reciprocal relationships with the analysis of the organic as well as historical mobility and dynamicity of the structure of the lexicon (Alinei, 1974, p. 10). In contrast to Coserius's position, Alinei pointed out that one of the problems of the compositional analysis is that of identifying a method that could

explain and be applied to the structure of the whole of the lexicon. To this regard, a compositional analysis of meaning, if applied to the all lexicon of a language, would allow to describe and retrace the *views of the world*[75] that are proper of a certain language both within a language and in useful inter-linguistic comparisons.[76]

Hence, the great theoretical innovation of Alinei's approach is that of having underlined the need of considering the structure of the lexicon as a whole and, consequently, of having highlighted the possibility of quantifying the semantic field as a portion of the whole lexicon. In order to understand the structure of a field and to establish the validity of the features that compose it, we need to reach all the linguistic units that share a certain feature and understand their position within the entire system. Lexical fields are characterized by an intertwined web of relationships within the lexicon as a whole: the same feature of meaning can appear in different sub-sets of the lexicon. For example, in the domain of the *horse*, *neigh* belongs both to the domain of the *horse* and of the sounds; *saddle* also belong to the domain of the artifacts and so on (Alinei, 1974, pp. 40–41). In Alinei's formulation of the principles underlying the componential analysis, we can find a clear criticism to the metaphor of the lexicon as a mosaic:

---

[75] According to Alinei, one important merit of the compositional approach is that of having combined the observations on the formal properties of the lexical systems with the most relevant results of the taxonomic research in biology and of having insisted on the important distinction between "taxonomy" and "nomenclature", i.e. on the difference between conceptual units and lexical units. On the one side, each distinctive feature is a level of opposition and hence, each definition of a lexeme can be represented in the format of a tree-graph in which the meaning of the word is given by the line the units the vertex of the tree to one of the points at the base of it. The distinctive feature that form such taxonomical structures are not lexical units, but rather concepts that are used to decompose the lexical unit. In this sense, language and thought do not overlap and the distinction between concepts and lexical units is of great importance. To conclude, this taxonomical structure is essentially hierarchically organized and constitute a formal universal of the human mind. It is the hierarchical order of the levels that determine the formation of the concepts; the content of the hierarchy can change with the evolution of society and culture but what does not change is the overall structure of thought (Alinei, 1974, pp. 20–25). We will not further develop the analysis of Alinei's view of the relationships between language and thought that, at least to us, seems quite obscure.

[76] The possibility of applying the compositional method and the field approach to cross-linguistic and cross-cultural studies is of great interest to us. Nonetheless, at this point we are mainly concerned with the validity of the methods as ways of describing and analyzing the structure of the lexicon within the framework of the research on meaning outlined in the previous chapter and, therefore, we will postpone the discussion about the possibility of applying it to cross-linguistic studies to later moments in our discussion. However, it seems useful to point out that, according to Alinei, the compositional analysis is a means by which we can identify and point out structural differences between closely-related as well as further away languages. The analysis of closely related languages allows Alinei to hypothesize a socio-psychological basis of the observed differences, given that they cannot be ascribed to important and deep cultural and environmental differences. In his example of the field of the sounds of the dog, he shows that Italian speakers tent to emphasize the more affectionate and whiny aspects, whereas Dutch and English speakers, although belonging to the same cultural environment, tend to emphasis less emotional aspects. These observations show that, even without drawing extreme relativistic conclusions, together with the undeniable effects of socio-cultural and psychological influences on the perception of the world, languages may impose forces that go in different or sometimes opposite directions (Alinei, 1974, pp. 40–41).

> L'idea che il lessico sia un tutto divisibile in sezioni, come un mosaico in pezzi combacianti l'uno con l'altro, è ancora oggi una delle più comuni, ma nello stesso tempo una delle più false e più dannose per una retta comprensione della natura e del funzionamento del lessico. [...] È come se si volesse classificare la popolazione italiana o di qualunque altro paese in base alle categorie delle "parentele": qui i "padre" là i "figli", da questa parte i "nonni" e da quella i "nipoti": dimenticando che ogni uomo può essere contemporaneamente padre, figlio, nonno, nipote, e ancora fratello, marito, cugino, zio, suocero, genero, ecc. pur rimanendo sempre "una sola" persona. In effetti, il lessico altro non è che una gigantesca "struttura di parentele" semantiche, in cui ogni membro riveste diverse funzioni pur restando sempre lo stesso (Alinei, 1974, p. 41).

Alinei explicitly excluded any theoretical debt with the Trier-Weisgerber theory of the lexical field inasmuch as it remained anchored to the literally interpretation of the Saussurian position and it did not "went through Prague" (Alinei, 1974, p. 21). However, it is clear that Alinei understood the lexicon as composed of a web of intertwined sub-sets that closely resembles a weak version of the notion of lexical field. Indeed, the definition of the boundaries (or better of the extension and quantity) of a certain lexical system is the prerequisite of any valid compositional description. Hence, the only way of verifying the completeness of a lexical field would be to examine the whole lexicon by means of a compositional analysis, which is clearly an impossible goal.

As a consequence of these theoretical observations, Alinei developed a method that would, at least in theory, allow him to reach all units of the lexicon that share a certain distinctive feature and to organize them into a coherent and comprehensive structure. Words were retrieved on the basis of an ingenious (and at the same time problematic) intuition: the words' definitions we find in traditional dictionaries are not very dissimilar from semantic-feature-based analysis. For example, the analysis of the verb *to neigh* provides the following compositional semantic features: SOUND and HORSE. Most dictionaries, in order to define the word *to neigh*, mention that it is a sound and that it is characteristic of a horse. If we generalize this observation to the whole of the entries in the dictionary, we end up with the working hypothesis that all lexical definitions in a traditional dictionary are an approximation of a rudimental formal definition we would obtain by means of semantic features analysis (cf. Alinei, 1974, p. 31 ff.). As one might imagine, this intuition is, on the one side, very productive because it can easily be integrated with computational methods that allow to examine a great amount of data; on the other side, it poses relevant theoretical and methodological issues that have been extensively addressed by the author in the mentioned work. Finally, by means of the implementation of the method we just mentioned, Alinei

obtains a list of around 300 words that share the semantic feature *horse*, i.e. these words have something to do with horses and are ready to be organized on the basis of their linguistic relationship. An in-depth analysis of the various steps and pitfalls of Alinei's methods and conclusions goes far beyond the goals of our work and, hence, we will only briefly mention some theoretical distinctions useful to understand how the author envisaged the structure of language.

As we have already said, according to Alinei, the lexicon cannot be compared to a mosaic. A theory of the structure of the lexicon has to explain the multiplicity of the relationships we can observe between words. In the first place, Alinei proposed to distinguish between *systems* and *domains*. *Lexical domains* are composed of lexemes that share at least one distinctive feature and that may belong to different grammatical categories and parts of speech, i.e. they are *asymmetric*; in the case of the domain of the horse, it would be composed by all lexemes that share the feature HORSE. *Lexical systems* are composed of lexemes that share at least one distinctive feature and that belong to the same grammatical category, i.e. they are *symmetric*; the system of the sounds of the horse is composed of lexemes that share the distinctive feature HORSE and belong to the same grammatical category. More specifically, a *lexical system* is organized around a basic relationship that assigns the functions of subject, object and of predicate to the two or three features of the system; for example, *abbaiare (to bark), ringhiare (to growl), ringhioso (growling),* and others are all part of a system because they share two distinctive features (DOG and SOUND) and because, for all these lexemes, DOG is the subject of SOUND. This basic relationship gives rise to different *lemmatic types* by means of logical-syntactic transformers; a *lemmatic type* is a kind of proposition starting from which each lexeme in the system takes its form and, hence, is the fundamental structure of the definition of its meaning. Alinei ends up with a list of 42 systems that compose the lexical domain of the *horse* and with an a even wider set of *lemmatic types*.

The method proposed by the author is overall quite complex and the results are considerably intricated; nonetheless, it seems worth noticing and underlining the effort to encompass in the analysis various aspects of linguistic meaning and especially the syntactic structure that cooperates in identifying the meaning of a certain lexeme. The structure of the lexicon does not depend entirely on the basic semantic features but also on the logical-syntactical transformations applicable to the basic semantic features.

At the same time, Alinei recognized the impossibility of reducing the lexicon to a finite set of basic and universal semantic features; in reducing the lexicon to domains, systems and

transformational rules, it becomes evident that language is fundamentally a dynamic system that cannot be compressed into a limited set of universals. The common error of those theoretical and methodological approaches that strive in the search for universal semantic primitives is that of considering language as an absolute and unique hierarchical structure. Conversely, in Alinei's view, the hierarchical structures by means of which languages are organized are intrinsically multiple, relative and functional. Semantic features are fundamentally socially and historically determined; in this sense, the semantic skeleton of the lexicon is grounded on the scientific, technical and ideological knowledge of a certain historical community and, thus, represents the whole of the human knowledge. This knowledge is not a set of unlinked and independent items, but rather a hierarchically structured compound of elements that is historically and socially determined (and hence subject to change) and that organizes our understanding of the world (Alinei, 1974, pp. 183–190).

3.4. How language became a cognitive phenomenon: cognitive turns and semantic primitives

In the previous paragraphs, we have mostly been dealing with some of the main developments in structural semantics when it comes to the definition and description of the meaning of words in historical-natural languages. To briefly summarize the main theoretical and methodological achievements of the above mentioned approaches, we could sketch the following scenario: the study of lexical semantics from around 1930 to around 1975 (cf. Geeraerts, 2006b, p. 387)[77] was dominated by the theoretical standpoint and methods of field and compositional approaches. The structure of the lexicon was mostly understood as independent and autonomous from the organization of extra-linguistic referents as well as from almost any aspect of cognition (of course with various degrees and with sometimes evident differences between the various authors). In those cases in which there was an attention for some kind of extra-linguistic facts, it was mostly directed towards the social and cultural components of meaning (recall, for example, Alinei's conclusion about the study of the field of the *horse* in Italian).

As we have already tried to point out, the above-mentioned structuralist approaches opened to fundamental questions and left many unresolved issues, such as the possibility of establishing the external as well as internal limits of a certain linguistic field, the problem of how to contrast the organization of a given linguistic field against the conceptual field that it should carve up, on what basis to compare different semantic fields over time (or across languages), the nature of semantic features, how they should be established and described, and so on. Many of these unsolved issues will be explicitly or implicitly taken on by cognitive approaches to language.

Let us clear out the aim of this and the following paragraphs: in the first place, we will try to give a general outline of the new questions posed by the birth and rise of cognitive linguistics; in the second place, we will more specifically turn to cognitive semantics and summarize some of the main developments within the field, which can be viewed as a continuation or a fracture with structuralist semantic approaches. We will inevitably come back to some of the issue discussed in chapters 1 and 2, as the view of language as a mapping device. The general aim of these paragraphs

[77] As pointed out by the author, this timeline is only an approximate and rough indication about the main period of dominance of a certain current of thought within the study of lexical semantics. As we will point our later (especially in paragraph 2.7), structural semantics methods had far reaching echoes in the following literature and the descriptive strategies of the componential and field approaches are still in use, even if revisited, in some of the current research paradigms. Nonetheless, in the last part of the last century, the attention in lexical semantics has progressively shifted towards a cognitive oriented type of research.

in to close the circle of our brief historical reconstruction of some of the leading views on the structure and organization of language. In order to do so, we need to start tackling the issue of language in cognition from its offspring.

3.4.1. *First* and *second* cognitive revolution: the broad picture

What do we mean by cognitive turn(s) in the study of language? What is cognitive linguistics and in what sense is it a cognitive approach to the study of language? In very general terms, cognitive linguistics is based on the theoretical standpoint that natural language has to be studied as a cognitive phenomenon.

The *first* cognitive revolution started in the 50s and is attached to the name of Noam Chomsky (e.g. Diodato, 2013, p. 103). Here, we are not interested in the specificities of Generative Grammar, but rather on the very general boost that the Chomskian paradigm, together with the rise of cognitive sciences more broadly, gave to the study of language in cognition. Chomsky turned the study of language inwards: language essentially became an innate and universal cognitive phenomenon. Chomsky *internalist* position declared an arrest of the behavioral view of meaning by setting as the focus of linguistic analysis the *competence* of the speaker as opposed to the *performance*, i.e. what any linguist should be interested in is not what the speakers *do* with language, but what they *know*. Rather than thinking about language as something that exists as such, as a system that can be studied in its own right, language received a mentalist interpretation. In the mainstream line of development of twentieth-century linguistics, the distinction between *competence* and *performance* is often interpreted in such a way that the *langue*-competence constitutes the proper object of enquiry of linguistics, whereas the *parole*-performance is of minor importance. Needless to say, this attitude has an affinity with the reductive views of lexical semantics that we have run into earlier: a strict separation between *langue* and *parole* in the field of grammatical theory resembles the strict separation of semantics and pragmatics in the field of lexical description (cf. Geeraerts, 2010).

Within the Chomskian paradigm, the relevant aspects of the faculty of language are its universal grammar and the transformational rules that allow it to operate. The faculty of language is a set of innate mental principles and rules that preside over language production and comprehension. Hence, within the first cognitive revolution, we can still see how the study of

language was essentially focused on a (revisited and internalized) level of the structure, leaving out all possible idiosyncratic uses (and leaving out meaning as well) (Chomsky, 1965). Language (i.e. the syntax of language) has to be studied as a cognitive, but still independent aspect of our cognitive life. In connection to modular views of the mind (J. A. Fodor, 1983), language could be seen as an autonomous module in cognition, responding to its own principles of functioning and organization. This first wave of cognitive linguistics is grounded on the very general intuition that the mind can be viewed as the *software* that runs the *hardware* (the body, from which it is independent and separated). Cognitive processes, including language, have to be analyzed as computational, algorithmic processes independent from the body (hence, this perspective is commonly addressed as *disembodied* cognition).

In the 70s, the modularity of the mind and the priority of syntax over semantics, central pillars of the Chomskian paradigm, are highly criticized by a group of young students coming from the Chomskian group, which initially gave rise to the generative semantics paradigm (we will talk more about this group later). The schism between semanticists and syntacticists is the starting point of the second cognitive revolution and of the forsaking of Chomsky's *disembodied* cognitivism. The *second* cognitive revolution is tightened to the birth of the paradigm of the *embodied* mind and arose as a critical rejection of the assumptions and methods developed by the first wave of cognitive linguistics. In the same period, the term cognitive linguistics (*cognitive grammar* and *cognitive semantics*) started to be used to describe a variety of different approaches that, beyond the many differences, share common assumptions on the nature of language and meaning. In the narrow sense, cognitive linguistics is one form of cognitive study of language that emphasizes the central role of semantics in the understanding of language.[78] From now on, the term *cognitive linguistics* will be used only in relation to the second cognitive revolution, as we will only be partially interested in approaches stemming from the Chomskian paradigm (and we will address them directly when needed).

After this terminological restriction, we are still faced with a wide variety of different and heterogeneous approaches that could fall under the umbrella of *cognitive linguistics* - some of the

[78] Geeraerts proposes to distinguish between a general (uncapitalized) cognitive linguistics that includes all approaches interested in the language as a cognitive phenomenon (and that includes generative grammar) and a Cognitive Linguistics that has as a major interest the study of meaning and, more precisely, of semantics within a cognitive frame of research (cf. Geeraerts, 2006a, p. 3). Although we endorse such terminological distinction, we will mostly focus on Cognitive Linguistics, and specifically on cognitive semantics, and hence we do not need to distinguish between the two, if not at this initial stage of our discussion.

most famous names in the field are those of George Lakoff, Ronald Langacker, Leonard Talmy, Ray Jackendoff, Charles Fillmore, Gilles Fauconnier, Philip Johnson-Lairs – whose interests and fields of research cover psychology, linguistics, artificial intelligence, philosophy of language, and more. Given this variety of approaches, in order to give a comprehensive account of cognitive linguistics, it will be necessary to scout for the more general underlying principles and assumptions that are shared from these approaches, leaving programmatically aside the sometimes-broad differences between the works of these (and other) authors.

As we have mentioned before, cognitive linguistics was born in a polemic contrast and rejection of some of the fundamental assumption of almost all previous paradigms of research interested in the nature and structure on linguistic meaning: namely, formalist, structuralist, and generative approaches.

Formal approaches to language are the main target of critique of cognitive linguists, which fundamentally refuse the idea that understanding the meaning of an utterance can be reduced to identifying its truth conditions. Cognitive semantics began as a reaction against the objectivist world-view assumed by the Anglo-American tradition in philosophy and the related approach, truth-conditional semantics, developed within formal linguistics. The main point of criticism to formal semantics is the exclusion of the lexicon and of the cognitive organization from a theory of meaning and semantics. Formal semantics is overall judged as psychologically implausible as well as inadequate in describing multiple aspects of natural languages (for an in-depth discussion of the theoretical opposition to formal semantics cf. Casadei, 2003b, pp. 43–44; Marconi, 1992).

On the same line of criticism, cognitive linguists openly reject many of the structuralist assumptions on the study of meaning[79]; they systematically reject to postulate and to employ theoretical devices without cognitive foundation, as, for example, the notion of invariant as structural principle of language (Luraghi & Gaeta, 2003, p. 23). Furthermore, structuralist and cognitive semantics set themselves on the opposite sides of the debate regarding two main lines of inquiry in the study of meaning (that we will extensively discuss later): the autonomy versus the non-autonomy of language (i.e. the dictionary view versus the encyclopedic view of meaning) and

[79] Taylor (1999) goes as far as recognizing the Saussurian heritage of cognitive theories of meaning in the notion of language as a symbolic system, fundamental for both theories, and in the idea that the meaning of an expression cannot be exhausted by the expression's designation, given that designation always takes place against a (more or less complex) network of background knowledge. In regard to the relationship between cognitive and structural semantics, it seems safe to conclude that, beyond perhaps some points of contact, many of the fundamental assumption vary drastically between the two theories.

the nature of the linguistic (as well as cognitive) categories. What is at stake with the second cognitive revolution is the fundamental structuralist claim that the lexicon of a natural language contains a proper semantic structure that is independent of the broader cognitive organization of the human mind, and which is the task of linguistic semantics to reveal. According to cognitive approaches, language and semantics have to be studied as a part of human cognition at large.

Definitely more complex to delineate is the relationship between generative linguistics and second-revolution cognitive linguistics. So much so that a proper discussion of it would definitely cover more than one chapter, and we will have to come back to it many times in the rest of our discussion. We will try to outline some of the main differences (and even some similarities) in section 3.5.1., when delineating some of the main principles and commitments underlying cognitive linguistics and, especially, cognitive semantics. At least at the current stage of our work, we will narrow down our discussion to position and theories interested mostly in the study of semantics, namely transformational semantics and cognitive semantics.[80] Before engaging in the presentation of cognitive view of meaning, let us briefly discuss some relevant issues at the intersection of generative theory and cognitive semantics. In the next two sections, we will first present Katz and Fodor's proposal at the intersection of compositional analysis and (broadly understood) cognitive linguistics. The main focus will be on the notion of *semantic primitives* as a consequence of translating compositional approaches according to the need of the first wave of cognitive approaches to language. In the second part, we will continue with the analysis of *semantic primitives*, presenting how this notion has been reframed in Wierzbicka's approach, namely in the framework of a *Natural Semantic Metalanguage*. As it might already be clear, our aim in not to discuss the generative/post-generative framework in the study of semantics. We will focus only on the specific issue of *semantic* and *cognitive* primitives, and we will try to outline their filiation from structural accounts of meaning as well as their most obvious shortcomings.

[80] We are well aware of the importance of the study of syntax for the rise of cognitive approaches to meaning. Although we will not address cognitive grammar as a specific phenomenon, some indications will be given throughout this chapter and a more specific elaboration will be presented when discussing usage-based approaches to meaning (Chapter 4).

3.4.2. Generativist semantics: the intersection of compositional analysis and cognitive linguistics

We left our discussion of compositional approaches to linguistic meaning with an open question about the nature of the semantic features involved in the description of the structure of the lexicon. As we have seen, in its original formulation, componential approaches did not entail any sort of cognitive realism of the distinctive traits. Nonetheless, because of the often-opaque distinction drawn by structuralist semanticists between the level of language and that of thought, the search for distinctive semantic traits left evident room for the transformation of the notion of *semantic traits* into that of *semantic primitives*.

Following from the Chomskian cognitive revolution and the consequent need to analyze language as a cognitive phenomenon, the Katzian model can be viewed as a combination of a structuralist method of analysis, a formalist system of description, and a mentalist conception of meaning. According to Geeraerts (2006b), generative semantics can be viewed as the culmination of (and at the same time the departure from) structuralist semantics: the Katzian definition of the methodology of lexical semantics within a generative framework is typically structuralist in that it assumes the existence of an autonomous level of semantic structure that is distinct from cognition at large. At the same time, the generative semanticists' tendency to broaden the empirical scope of generative grammar contained a movement away from the autonomy aspired by Chomskian linguistics. If analyzed from the point of view of the intersection between structuralist and cognitive accounts of meaning, this approach contributed to raising important questions about the cognitive reality of semantic descriptions and, specifically, about the cognitive adequacy of compositional accounts of meaning.

To this regard, it seems important to note that structuralist compositional approaches were in general not concerned with the psychological validity of their thesis, but rather aimed at developing a description of language within language; hence, structuralist compositional approaches were far from theoretically stating the ontological reality of their descriptive components and, if understood as such, they may still show their potentials in describing the structural organization of language (cf. Diodato, 2015, p. 23). When the componential approach was merged with the generative framework in the work of Katz and Fodor, the object of investigation shifted from the structure of language (in a structuralist sense) to the query about the ability of the language users to interpret sentences (Katz & Fodor, 1963, p. 176). Hence, generative

semantics diverged from traditional structuralist approaches to the extent that it was not interested as much in the organization of language, but rather in the way in which we understand language. For instance, Katz and Fodor's componential analysis does not take as its starting point the contrastive analysis of a set of words belonging to the same lexical field, but rather gives an example of the way in which the different meanings of one single word, when analyzed componentially, can be represented in a formalized dictionary as part of a formal grammar allowing to disambiguate sentences on the basis of a (somewhat revisited) componential approach.[81]

At the same time, Katz and Fodor's approach set the ground for the future scission of cognitive linguistics from the generative paradigm by focusing on the importance of lexical semantics for language comprehension and description. In the initial phases of generative linguistics (Chomsky, 1957), no place was reserved for semantics. It was mostly thanks to Katz and Fodor's work that a semantic description was incorporated into the generative framework. As a result of operating at the conjunction between lexical semantics and formal grammar, the merit of Katz's approach was the incorporation of syntagmatic relations to the set of phenomena under scrutiny in formal grammar, which had to account for both paradigmatic and syntagmatic relationships and to show how both could follow automatically from the underlying featural representations of meaning and the operations of projection rules.

Although remaining substantially loyal to the structuralist need to distinguish between semantic and encyclopedic knowledge[82], the conjunction of cognitive theoretical claims and

[81] In this concise presentation and discussion of Katz and Fodor's approach, we are not specifically interested in the specificity of their componential analysis but rather on the theoretical implications of the bond between componential and cognitive accounts of meaning. Nevertheless, it is worth mentioning some of the notions introduced by the authors to better grasp their conception of language and their final goals. We can start from the well-known example of the componential analysis of the English word *bachelor*. Next to word form and word class, two types of semantic component are identified: markers, which constitute the 'systematic' part of the meaning of an item (the syntagmatic or selection restrictions), and distinguisher, which aim at representing what is idiosyncratic about the meaning of an item. The formal mechanism to disambiguate sentences on the exclusion of semantic anomalies consists in the so-called projection rules which are responsible for the combination of lexical meanings of individual words in a sentence into constituent meanings (for a more detailed explanation cf. for example Geeraerts, 2010, pp. 103–106).

[82] Katz and Fodor set the scope of the semantic endeavor in the description the ability of the users to interpret sentences and thereby recognize that the act of interpretation involves the full extent of the knowledge of the user not just about language but also about the world. At the same time, in order to narrow down the aim of a theory of semantics, the authors set as "upper limit" of the theory the need to focus on knowledge of language, and not on knowledge of the world; "grammars seek to describe the structure of a sentence in isolation from its possible settings in linguistic discourse (written or verbal), or in non-linguistic context (social or physical)" (Katz & Fodor, 1963, p. 173). For a discussion of the issues faced by the authors in dealing with the relationship between sematic and encyclopedic knowledge cf. Geeraerts (2010) and Violi (1997).

componential methodology led to a substantial change in the status of semantic features: semantic features are no longer only necessary and sufficient semantic components but turned into a closed and limited set of semantic primitives with a conceptual status. Semantic features like HUMAN, MALE or ADULT constitute the concept of *man,* rather than just specifying the semantic features of the lexeme (cf. Diodato, 2015, p. 20). Katz and Fodor's approach suggested that the lexical entries of human languages can be analyzed, partially or completely, by means of a limited set of semantic components that are the same for every language, independently from the their particular semantic structure; semantic components are introduced in the semantic theory to designate invariant linguistic behaviors, connected to language, of a conceptual system that is part of the cognitive structure of the human mind (cf. Katz, 1966). It is almost needless to point out the evident pitfalls that follow from the hypothesis of a limited set of atomic, universal, and trans-cultural concepts as part of the general faculty of language. In the words of Lyons,

> little need be said about the alleged universality of semantic components, except that it is an assumption which is commonly made by philosophers and linguists on the basis of their anecdotal discussion of a few well-chosen examples from a handful of the world's languages. […] It may well be that future developments in semantics, psychology, physiology, sociology, anthropology, and various other disciplines, will justify the view that there are certain 'language invariant […]'. Such empirical evidence as there is available at the present time would tend to refute, rather than confirm, this hypothesis (Lyons, 1968, p. 473).

Despite the evident difficulties in endorsing the cognitive nature on semantic primitives, one of the most successful approaches to date is represented by the attempt made by Miller and Johnson-Laird to anchor semantic primitives onto *perceptual primitives*: according to the authors, the majority of the concepts are extracted from our perception of the world and connected to it (Miller & Johnson-Laird, 1976; cf. Violi, 1997, pp. 110–112). Perceptual primitives, on which semantic primitives are based, are well-suited to fit the general framework of a cognitive oriented semantics, as they directly link language to our perception of the world. However, *stricto sensu*, they cannot be understood as the reductional atoms of our conceptual representations; perceptual primitives are essentially described at the interface with perceptual experience and, therefore, are accessible only by means of their ostensive relationship to the world (cf. Diodato, 2015, p. 21). On the contrary, according to the generative paradigm, semantic primitives have to explain the formation of more complex concepts and, hence, should be understood as conceptual in nature.

3.4.3. Semantic primitives in *Natural Semantic Metalanguage*

In continuing with our discussion of semantic primitive, Wierzbicka's proposal of a *universal Natural Semantic Metalanguage* extensively discussed the cognitive foundations of componential descriptions of meaning in the attempt to come to terms with the dilemma of a cognitive turn in lexical semantics, i.e. to reconcile the reductionist tendency of componential analysis with the need to take cognition seriously. The main problem that arises in the combination of cognitive linguistics and compositional approaches is outlined by Geeraerts in these terms:

> on the one hand, a decompositional method has a reductionist tendency: it reduces the semantic description to a set of primitive meaning components, and looks for a truly linguistic level of description, contrasting with an encyclopedic level. On the other hand, a theory that aims at psychological adequacy will inevitably have to face the fuzziness and flexibility of language use (Geeraerts, 2010, p. 126).

Wierzbicka abandoned the idea that meaning components derive from distinctive oppositions within the lexical field, but rather assumed the existence of a universal set of semantic primitives that may be discovered by defining words through a process of reductive paraphrase that would lead to the discovery of universal *semantic primes* lexicalized in all languages of the world.[83]

The actual universality of the so-discovered primitives is still a matter of debate. The list of semantic primitives suggested by Wierzbicka and collaborators has changed and increased in number over time, starting with 14 primitives (cf. Wierzbicka, 1972) to now 50 (cf. Wierzbicka, 1996) with the consequence of decreasing the intuitiveness and immediacy of their universality (cf. Violi, 1997, pp. 106–107). Even more complicated is the issue of how to assess the supposedly universality of the primitives. In fact, not only the reductive method does not allow to identify which set of primitives derived from the reduction might be more adequate compared to others, leaving room for many different basic vocabularies, but, at the same time, the idea that a definition that makes use of true primitives should be more intelligible to speakers that any other has been various disproved. The intelligibility of a definition probably depends less on whether its

[83] The 'reductive paraphrase' method consists in writing definitions couched in the vocabulary of universal primitive concepts. In the actual practice, the definition often makes use of non-primitive elements that then need to be further analyzed to reach the level of the primitives.

components are conceptually primitive than on the question whether they are previously known and familiar to the reader (cf. Riemer, 2006).

Another fundamental assumption underlying the search for semantic primes is that "the mind is neat but the world is fuzzy" (Geeraerts, 2010, p. 127). Wierzbicka explicitly defends the discreteness of semantics above and beyond the vagueness of the referents for a given word. In order to define the meaning of a word, it is necessary to study the structure of the concept which underlies and explains its applicability: the information included in the definition does not primarily involve the objective features of the referents of the expression, but rather what people think about those referents (cf. Wierzbicka, 1985). The referential use of a word may be flexible, heterogeneous and not clearly delineated; on the contrary, the concept that is realized in all of those instances of use is precise, discrete, and uniform. Hence, the variability is the basis for a reductive strategy that has the aim of describing the semantic invariant that is present in all instances of variation. According to this method, if the definitions are supposed to be generally applicable to all instances in which a word is used, this fact should be empirically validated on the basis of a broad sample of actual language use to check whether the features that are regarded as universals within a concept indeed apply to all the relevant cases (for empirical tests cf. Geeraerts, 1993; cf. Geeraerts, Grondelaers, & Bakema, 1994). Not even the intuitiveness of the definitions seems to hold in the face of the introspection and intuitions of the speakers; the mechanism of introspection is itself not very likely, since "it seems to imply that the actual context of use would not be able to override the 'essential' features residing in the concept. So, even if the word cup is applied to an extremely peripheral instance of the category (say, a plastic cup in your desk drawer filled with paperclips and pencils), you would still think of that cup as a drinking vessel, and if fruit is used to talk about strawberries, you would still think of it as having a skin that is harder than the parts, or as growing on trees" (Geeraerts, 2010, p. 137) or, as Kay remarks, "it does not seem to me that when I say that a particular car, say, is green what I have in mind is that when I see things similar to this car I can think about things that grow out of the ground" (Kay, 2004, p. 240) as Wierzbicka's definition of the word *green* would imply. More broadly,

> Da un lato infatti è chiaro, soprattutto nei lavori più recenti di Wierzbicka, che l'oggetto della sua ricerca è l'inventario dei "concetti umani fondamentali", per usare le sue parole, che quanto più sono universali e innati, tanto più dovranno appartenere a un livello molto profondo e generalissimo di descrizione. Si tratta di invarianti che potrebbero essere più propriamente definite come universali dell'immaginario umano,

> schemi archetipici di fondo che hanno a che vedere con una fenomenologia corporea di base, con le disposizioni affettive elementari dell'essere umano, con categorie molto generali dell'esperienza. […] Anche se certamente la struttura dei significati linguistici reca una qualche traccia di queste salienze profonde, ed è in questo senso largamente attraversata da un'istanza di motivazione extralinguistica, non è certo ipotizzabile alcuna immediata e diretta derivabilità dei singoli significati linguistici dall'insieme degli universali profondi (Violi, 1997, p. 109).

The problem lies in the overall unjustified transformation of the "fundamental human concepts" into semantic primitives from which to derive all possible linguistic determinations in all possible languages. Semantic compositional approaches rest on the hypothesis that semantic knowledge is (at least at some level) innately pre-specified. From this perspective, while languages may differ "on the surface", beneath the surface they are broadly similar, and this similarity is explained by the existence of a set of primitives (together with a universal set of principles that operate on those primitives in the case of generativist semantic approaches).

Furthermore, the traditional view in semantic theory holds that meaning can be divided into a dictionary component and an encyclopedic component. According to this view, it is only the dictionary component that properly constitutes the study of lexical semantics - the branch of semantics concerned with the study of word meaning. In contrast, encyclopedic knowledge is external to linguistic knowledge, falling within the domain of "world knowledge". From this perspective, then, dictionary knowledge relates to knowing what words mean, and this knowledge represents a specialized component, the *mental dictionary* or *lexicon* (theories of the mental lexicon will be the main focus of the next chapter).

According to the perspective adopted in second-generation cognitive semantics, the strict separation of lexical knowledge from "world knowledge" is problematic in a number of ways. To begin with, the dictionary view assumes that word meanings have a semantic "core", the "essential" aspect of a word's meaning. This semantic core is distinguished from other non-essential aspects of the word's meaning, such as the associations that a word brings with it. The dictionary view assumes that words, although related to other words by lexical relations like synonymy and so on, can nevertheless be defined in a context-independent way. In contrast, as we will extensively discuss in the following sections, a number of scholars, as for example Fillmore (1975, 1977, 1982; and Fillmore & Atkins, 1992) and Langacker (1987) have presented persuasive arguments suggesting that words in human language are never represented independently of

context. Second-generation cognitive linguists argue that words are always understood with respect to frames or domains of experience. Parallel to the attempt of maintaining semantic knowledge and world knowledge separated, compositional accounts of meaning within a cognitive framework imply, at least to some extent, to assume a more or less static view of conceptual and linguistic categorization that will be fundamentally challenged by prototypical accounts of the dynamical structure of conceptual categories. In the next section, we will hence move to the analysis of cognitive semantics approaches, arguably the most popular framework for the study of meaning in contemporary linguistics.

3.5. Cognitive semantics: foundational principles and language structure

3.5.1. Commitments

As we have mentioned before, second-generation cognitivists depart from many of the intuitions and assumptions proper of the first cognitive revolution. We have already hinted at some of the main differences in the previous sections. Here, we will offer a summary of some of the most relevant programmatic assumptions of second-generation cognitive approaches to language. It should be noted upfront that not all approaches in second-generation cognitive linguistics share the same interests, methods, and theoretical positions. However, despite the often-profound differences, they all seem to share some very general commitments and intuitions.

A general principle of cognitive approaches to meaning is what Lakoff phrased in terms of *cognitive commitment*: the idea that a model of language can be a cognitive one only if it goes along with and is supported by empirical data on the functioning of the mind (Lakoff, 1990, p. 40). If we take this perspective as a general claim about the kind of analysis put forward by cognitive linguists, to work in the frame of cognitive linguistic means to endorse the need for the cognitive plausibility of the theoretical models, which, in turn, points to the general assumption that, as we will see (and as we have partially already seen in Chapter 1), language and cognition cannot be taken as independent one from the other (and, especially, language cannot be understood independently from cognition).[84] The cognitive commitment leads to the view that principles of linguistic organization should reflect general cognitive principles, rather than cognitive principles that are specific to language.

The second commitment is that there are common structuring principles that hold across different aspects of language, and that an important function of linguistics is to identify these principles. Again, this sets second-generation cognitive linguists in contrast to the Chomskian assumptions about the nature of language and the way to study it. Generativist accounts of language generally take phonology, semantics, and syntax as concerning significantly different kinds of structuring principles operating over different kinds of primitives. For instance, a syntax "module" would be an area in the mind concerned with structuring words into sentences, whereas

[84] Whether cognitive theories of meaning actually are cognitive plausible or not is still a matter of an open debate, which goes beyond our scopes and need. For an introduction and discussion of this issue we refer to Casadei (2003b, pp. 52–53).

a phonology "module" would be concerned with structuring sounds into patterns permitted by the rules of any given language, and by human language in general. Cognitive linguistics acknowledges that it may often be useful, for practical purposes, to treat areas such as syntax, semantics, and phonology as being notionally distinct, but rejects the idea that the "modules" or "subsystems" of language are organized in significantly divergent ways, or indeed that distinct modules or subsystems even exist. Different language components can be seen to share fundamental organizational principles.

Taken together, these two general trademarks of cognitive approaches to language bring to the rejection of the modularist view of the mind and of the independence (and priority) of grammar over other aspects of language. In the development of cognitive linguistics, we observe a progressive drift away from the lead role of grammar towards an approach fundamentally grounded in the idea that meaning has to be the main focus of interest. Cognitive linguists criticize the generativistic idea that grammar and syntax are constitutive of language in a way that is autonomous from meaning; one of the main goals of cognitive semanticists is to rejoin these two aspects of language (study of meaning and study of syntax) by focusing on the primary role of meaning as the structural principle of language. The grammar and the lexicon of a language cannot be taken as two clear-cut separated levels: grammatical forms themselves have meaning which is not essentially different from the meaning of lexical units. *Cognitive Grammar* (formulated by Langacker, 1987) and *grammatical construal* (a perspective generally linked to the work of Talmy, 2000) are clear examples of theories of grammar that focus on meaning, and in which the typical formal categories of grammatical descriptions are reinterpreted from a semantic point of view.

Assuming that we buy into these three main arguments (cognitive and generalization commitments and interdependency of syntax and semantics), what can we say about the nature of language and of linguistic meaning?

In the first place, language is about comprehension. As stated in the Editorial Statement of the first issue of the Journal of Cognitive Linguistics (1990), language is an instrument for organizing, processing and conveying information; the processing and storage of information is a crucial design feature of language that itself is a form of knowledge (cf. Geeraerts, 2006a). Hence, the link between language and knowledge, between semantics and comprehension, is a fundamental assumption of cognitive accounts of language. To describe meaning means to describe the way in which we understand linguistic expressions. As Fillmore observes, given that

comprehension is a mental activity, cognitive semantics needs to describe meanings as cognitive contents, i.e. as what the mind makes use of to understand a linguistic expression (cf. Violi, 1997).

We have already encountered the idea that meaning is a matter of cognitive content in the previous chapters. Here, we will go a little more into the details of the nature of meaning in cognitive linguistics tackling onto the notions of encyclopedic meaning and *embodied* meaning.

The notion of *embodiment* (or of *embodied cognition*) is a central idea in cognitive linguistics. Cognitive linguists emphasize the importance of human experience, the centrality of the human body, and of the human-specific cognitive structure and organization, all of which affect the nature of our experience. The human mind – and therefore language – cannot be investigated in isolation from the human body. The idea that experience is embodied entails that our construal of reality is likely to be mediated in large measure by the nature of our bodies. The nature of our biological morphology, together with the nature of the physical environment with which we interact, determine many aspects of our experience. The fact that our experience is embodied has consequences for cognition. The concepts we have access to and the nature of the "reality" we think and talk about are a function of our embodiment. The human mind bears the imprint of embodied experiences. In his now classic book, *The Body in the Mind* (1987), Mark Johnson proposes that one way in which embodied experience manifests itself at the cognitive level is in terms of *image schemas*. These are rudimentary concepts like CONTACT, CONTAINER and BALANCE, which are meaningful because they derive from and are linked to our pre-conceptual experience. Lakoff (1987, 1990) and Johnson (1987) have argued that embodied concepts of this kind can be systematically extended to provide more abstract concepts and conceptual domains with structure. In the *embodied* perspective put forwards by Lakoff and Johnson (1999), cognitive linguistics regains the focus on the intrinsic connection between language and experience: very broadly understood, meaning is experientially grounded in our sensorimotor experiences.[85]

[85] Since the establishment of the Chomskyan mentalist model of language in the mid-twentieth century, which firmly focused attention of researchers on language as a cognitive phenomenon, and the simultaneous rise of cognitive science, theories of mental representation have adopted a non-perceptual view. This is sometimes called the *amodal* view, because it does not view conceptual structure as based not on perceptual (modal) states, but rather on a distinct kind of representational system. According to Barsalou, cognitive sciences were influenced in this respect by formalisms that emerged from branches of philosophy and mathematics (such as logic), and from the development of computer languages in computer science and artificial intelligence. Moreover, the prevalence of the modular theory of mind, not only in linguistics but also in cognitive psychology, represented a widespread view of perception and cognition as separable systems, operating according to different principles. As we have seen, recent developments in cognitive psychology suggest that the conceptual structure actually has a perceptual basis (it is inherently *modal*).

A fundamental concern for cognitive semanticists is the nature of the relationship between conceptual structure and the external world made available to us through our sensory experience. Embodied cognition theses hold that the nature of our conceptual organization arises from bodily experience, so part of what makes conceptual structure meaningful in the bodily experience with which it is associated. As we have seen in chapter 2 (see especially paragraphs 2.1.2 and 2.1.3), this leads to the view that semantic structure can be equated with concepts. Linguistic meaning is conceptual, and concepts are grounded in our sensorimotor experience of the world. In the classical cognitive view endorsed by generativism, thought was viewed as a manipulation of abstract symbols and the functioning of the mind was independent from the physical characteristics of the human body, according to the principle that the software is independent from the hardware. In contrast, second-generation cognitive scientists assume no separation between the mind and the body and presume that the core of our conceptual system originates precisely from the bodily experience we have of the external world. The experience we have of the world is far from being completely chaotic and, instead, it is largely organized and structured, previous to the intervention of conceptualization, into basic pre-conceptual schema that constitute the foundation of our conceptual and semantic-conceptual system (an example could be Johnson's notion of *image schema*). In this sense, meaning in not separated from the way in which we interact with the external world, from other forms of knowledge, and from the way in which these various forms of knowledge are integrated with other cognitive abilities. As we have already begun to see, word meanings, and hence – within this perspective - concepts, cannot straightforwardly be defined. For this reason, cognitive semanticists reject the definitional or dictionary view of word meaning in favor of an encyclopedic view. This means that words do not represent neatly packaged bundles of meaning (the dictionary view) but serve as 'points of access' to vast repositories of knowledge relating to a particular concept or conceptual domain (e.g. Langacker, 1987). Lexical concepts (the portion of conceptual meaning encoded in language) can only be studied adequately against the background of the human cognitive capacities at large and, in particular, there is no specific linguistic, semantic organization of knowledge separated from conceptual memory in the broadest sense.[86]

These ideas, together with the empirical evidence that is beginning to be gathered, is consonant with the claims of cognitive semantics, and particularly the thesis of embodied cognition.

[86] In chapter 2, we have already mentioned something about the view that linguistic meaning is essentially conceptual. This brief introduction on the encyclopedic and embodied nature of meaning reinforces the idea that, within the

Before turning to some specific instances of this general principles, allow us some comprehensive remarks about the novel insights that a cognitive perspective on meaning brings to the table. Linguistic meaning is dynamical and flexible; new experiences and changes in our environment require that we adapt our semantic categories to transformations and circumstances, and that we leave room for nuances and slightly deviant cases. Hence, language cannot be a more or less rigid and stable structure, but should rather be thought of a continuous process of signification and meaning construction, of conceptualization, in a constant recruitment of background encyclopedic knowledge. The fundamental process at play is the way in which we segment our experience of the external reality by distinguishing between different categories. This constitutes the foundation not only of our construction of concepts but, at the same time, of the possibility of using a symbolic system such as language; every linguistic sign is a categorization and pertinentization activity. This theoretical consideration of the dynamicity and flexibility of meaning will lead, together with the encyclopedic view of meaning, to the need of reconsidering the structure of conceptual (as well as linguistic) categories. Although cognitive linguistics is not a completely new approach in the history of linguistic ideas[87], it has the great merit of having outlined a theoretical framework able to integrate in a unified perspective the need to consider both the mental and the extra-linguistic levels as constitutive of meaning, and the vague nature of the linguistic signs as a fundamental aspect of both our linguistic practice and of the linguistic inquiry. Starting from the reconsideration of the structure and process of categorization, up to more comprehensive accounts of the possible structure of our experiential-conceptual-semantic knowledge of the world, cognitive semantics openly deals with the intrinsic multifaceted nature of our experience.

approaches described here, language equals concepts. We have already shown some of the possible shortcomings of this view in the previous chapter. In chapter 5, we will come back to notions of meaning as concepts and of meaning as conceptualization, offering a somewhat different perspective on the issue. We are aware that we are coming back to this topic over and over, dealing with it from a slightly different perspective every time.

[87] Various attempts have been made to individuate the theoretical ancestors of cognitive semantics approaches and especially regarding the non-autonomous view of meaning; for example, Geeraerts (2006b) points to the pre-structuralist psychological orientation of Breal's thesis on the nature of language. For other historical perspectives see Simone (1992) and Casadei (2003b).

3.5.2. *Prototypes* as categories

One key aspect of human cognitive activity is categorization, which, in a very simplified way, "involves the apprehension of some individual entity, some particular of experience, as an instance of something conceived more abstractly that also encompasses other actual and potential instantiations" (Croft & Cruse, 2004, p. 74).

As we have already seen, cognitive linguistics generally assumes that language reflects conceptual structure and organization. Needless to say, then, within the cognitive linguistics perspective, to study thought means to deal with the problem of categorization; given the "cognitive commitment" and the centrality of categorization for human thought, we expect this ability to be reflected in linguistic organization. Both cognitive psychologists and semanticists require a theory of categorization to account for knowledge representation and, hence, for linguistic meaning. The idea that language implies some form of categorization is definitely not new in the history of linguistic idea; as we have pointed out in the previous chapters, the Saussurian view of meaning, as re-analyzed in De Mauro's approach, as well as relativistic accounts of meaning, firmly suggest that language is fundamentally a categorization and pertinantization device.

When compared to structuralist approaches to meaning, the fundamental innovation brought together by cognitive approaches regards the structure of the categories. Componential accounts of meaning, both in structuralist and in cognitive semantics,[88] were fundamentally based on the classical – otherwise called Aristotelic - view of categories; the so-called classical model of categories defines them in terms of a set of necessary and sufficient features. The features are necessary in that no entity that does not possess the full set is a member of the category, and they are sufficient in that possession of all the features guarantees membership. The classical model established clear and rigid boundaries to a category.

[88] It seems relevant to stress again the fact that structuralist approaches, including componential approaches within a structuralist framework, never claimed (and were theoretically very far from claiming) that their description of linguistic categories was, at the same time, a model of conceptual categories. Thus, structuralist approaches to meaning cannot and should not be criticized on the basis of their cognitive implausibility, since they were not a model of cognition. The situation is different for those cognitive approaches that are grounded on the traditional model of categories; in fact, if the set of features that constructs the category is, by definition, a cognitive primitive, then it should stand the proof of cognitive plausibility.

Some frequently cited shortcomings of this model have provided the major motivation for the development of alternative theories. Classical theories do not seem to offer explanations for why many concepts cannot be adequately defined in terms of necessary and sufficient features; for the fact that many definitions hold only within a specific domain; for the fact that not all members of the categories have equal membership, in the sense that some seem to be more representative of the category; for why category boundaries seem often vague and variable (fuzzy).[89]

Furthermore, the widespread phenomenon of polysemy in language seems hard to explain in the framework of a traditional view of categories; if we look at cluster of senses associated with the same lexical entry, it is easy to note how fully-fledged dictionary definitions could hardly capture it. How do these different senses connect? Are there cognitive mechanisms that would allow us to explain this phenomenon? Can we systematically distinguish between the semantic level (that of the senses) and the referential level (that of category members)?[90]

If we look at how cognitive linguists and psychologists tried to solve these issues, we come across a new idea that will become the ground for many research paradigms: linguistic categories may be fuzzy at the edges but clean in the center. This is the idea behind prototypical accounts of conceptual as well as of linguistic categories.

One of the first psycholinguistic researches to have shown the inadequacy of a classical view of categories is that of Berlin and Kay (1969) on basic color terms; the authors were able to show that, both intra-linguistically as well as cross-linguistically, speakers tend to select the same 'best example' of a certain color whereas wide differences are recognized in the delimitation of the boundaries of each term, suggesting that, independently from the language spoken, people tend to agree on which colors are focal and on the best members of a certain color category.

Next, Labov (1973, 1978) studied the variable denotation of cups and cup-like containers, i.e. the fact that words like *cup, mug, bowl* or *glass* may refer to objects that do not have a uniform appearance. The conclusion is that a word like *cup* is referentially vague: the exact boundary of

[89] A classic example is provided by observing the structure and definition of the category *bird*: is a chicken a *bird* even if it cannot fly, if *to fly* is a necessary feature of the category? Why is a robin 'more' a *bird* than a penguin? If we remove those features that are not shared by all *birds*, are the remaining features sufficient to distinguish birds from other species? (cf. Geeraerts, 2010). For an useful discussion of the history and critics of traditional category theory see, among others, Casadei (2003a, pp. 92–95).

[90] A separate work would be required to properly address the issue concerning the distinction between referential vagueness and polysemy and between the various ways in which we can say that language is indeterminate. For an in-depth discussion see Geeraerts (1993, 2010, pp. 196–199).

the denotational range of the word is not immediately given. According to Labov, lexical categories have fuzzy edges; the membership to a category is not determined by the basic attributes or features of the category, but rather it depends upon functional, interactional and cultural aspects.

The psychologist Eleanor Heider Rosch tested the hypothesis put forward by Berlin and Kay and developed a full-blown theory based on two fundamental assumption: firstly, non-linguistic factors, such as perceptual ones, have a crucial role in the creation and structuring of categories; secondly, categories are organized around an informative center constituted by the best representative of the category, the prototype. As for the structure of the categories, Rosch's studies led to the conclusion that the tendency to define categories in a rigid way clashes with actual psychological data: perceptually based categories, like colors, do not exhibit sharply delimited borders. Categories are unambiguously defined in their focal points. On the contrary, the areas at the margins of the categories do not exhibit a clear-cut definition. The theory implies that the range of application of such categories is concentrated around focal points represented by a prototypical member of the category. The attributes of these focal members are structurally the most salient properties of the concept in question, and conversely, a particular member of the category occupies a focal position because it exhibits the most salient features.[91] The reason behind such notion of category is that the process of categorization is not arbitrary, but rather guided by the cognitive need to optimizing the information from the external world. "Categories have to respond to a principle of cognitive economy that consists in giving the maximum information with the minimum effort, and this function is reached by gathering together entities that clearly resemble each other and, at the same time, clearly differ from other entities" ([our translation] Casadei, 2003a, p. 97). In order to define whether a certain entity belongs to a category we do not need to

[91] Rosch proposes a new account of the internal structure of the category both on the horizontal dimension and on the vertical dimension of hierarchical relationships. On the horizontal axis we find categories at the same level (*chairs, table*, ecc.), each included in superordinate category (*furniture*) and a subordinate category (*deck-chair*). On both the vertical and the horizontal axes, there is a cognitively more salient area: on the vertical axis, the most salient area is that of the basic level of intermediate categories such as *chair*. Similarly, on the vertical axis, not all members have the same status; *robins* are more *birds* than *penguins*, *chairs* are more *furniture* than *telephones* and so on (the prototype of the category). For an in-depth discussion on the structure of categories and on the privileged cognitive status of the basic level see Cruse and Croft (2004) and Diodato (2015). Prototypical accounts of category consider multiple dimensions: on the one side, they take into account the referential, extensional structure of a category by showing that not all members of the category are equal in the representativeness for that category and that the referential boundaries of a category are not always determinate (non-equality and non-discreteness). These two properties recur on the intesional level, i.e. in the definitional structure of the category; non-equality involves the fact that categories have internal structure: not all members or readings that fall within the boundaries of the category need have equal status, but some may be more central than others. Categories often consist of a dominant core area surrounded by a less salient periphery (cf. Geeraerts, 2010, pp. 188–189).

check a list of specific properties, but rather we need to verify how similar or dissimilar the entity is from the prototype.

Rosch's prototype results were introduced in linguistics in the early 1980s and, in the course of the linguistic elaboration of the model, it became clear that it was important to distinguish between the various phenomena that may be associated with prototypicality. It should be noted that the theory of prototypes was born to explain how we categorize and conceptualize reality and not to identify the mechanisms of semantic categorization. Nonetheless, Rosch's theory had a great development within cognitive linguistics. In turn, this intersection brought to the recognition of some of its shortcomings.

How can we use the notion of prototype to describe meaning in language? As Casadei clearly shows, it seems difficult to directly apply the classical interpretation of prototypes - the prototype is the best representative of a certain category - to semantic analyses. It would be difficult to endorse a theory of meaning according to which bird means *robin* or the typical meanings of furniture are *chair* and *table*. A revisited version of the prototype theory proposes to understand the notion of prototype not as a concrete, but rather as an abstract entity: the prototype is an abstract construction, a conceptual schema, that unifies all typical properties of a category. For example, the meaning of bird would be represented by the intersection of "knows how to fly", "has wings", "has a beak", etc., which, in turn, find their best actualization in the robin (cf. Casadei, 2003a, p. 99). Despite the efficiency of moving the notion of prototype to the level of abstract and mental representation to account for linguistic meaning, this move poses the problem of a gap between the linguistic and the psychological level; for instance, psychological data show that judgments of prototypicality are based on the referential value more than on abstract concepts, i.e. that when we think of a bird we tend to form a mental image that corresponds to a central representative of the category.

Despite this and other related problems, which are intrinsic to the notion of prototype, the theory remains appealing for explaining semantic structure inasmuch as it based on the fact that categories boundaries are not clear-cut boundaries, but rather malleable and fuzzy edges. Prototype theory seems apt to represent one of the most salient properties of linguistic meaning, namely vagueness (or at least referential vagueness). Nonetheless, the notion of prototype does not overlap with the notion of vagueness, since categorical representativeness and category membership are not identical phenomenon. For example, penguins are birds: we have no doubts about their

memberships in the category *bird*, even if they are not the highest representatives of the category, i.e. they are far from the prototype. Hence, the notion of prototype turns out to be insufficient to explain the structure of the category, and our judgments about the membership of an item to a category does not seem to necessarily depend on the prototypical structure of the category (cf. Casadei, 2003a; Diodato, 2013, 2015).

Ultimately, do prototypes allow us to distinguish and explain *referential polysemy* (the fact that *bird* refers to a variety of denotata) and *semantic polysemy* (the fact that *school* has various senses, like "the school building" or "a school of thought", and hence it exhibits multiple acceptations)? In order to account for how different categories exhibit different structures, and for the distinction between non-discreteness and non-equality of categories, Geeraerts (1989, 2010, p. 189) proposes to consider the notion of prototype as a prototypical notion itself. Hence, a category can be prototypical to some extent: for example, polysemous categories, although including more or less representative entities in the category, are not organized around one prototypical member. Polysemous lexemes encompass a variety of senses and acceptations and, therefore, cannot be organized around one single prototype but, at least, should have one prototype for each sense encompassed by the lexeme (for a discussion see Casadei, 2003a, pp. 104–108).

In conclusion, prototype semantics offers a valid alternative to classic models allowing to overcome some of their standard issue (e.g. the clear-cut distinction between dictionary and encyclopedia and the representation of meaning in the form of a check-list) by theoretically connecting the study of meaning to the cognitive as well as referential dimensions. Drawing from the consideration of the various shortcomings in effectively explaining vagueness and polysemy in language, we can see how the theory of prototypes faces a general issue (which is shared with other cognitive accounts of meaning as well): it flattens linguistic meaning onto concepts and onto concepts organization, leaving out many substantial linguistic phenomena. In structuralist accounts of meaning, the problem was that they had isolated linguistic phenomena from the cognitive as well as referential world and, consequently, offered only a partial account of linguistic meaning. Conversely, cognitive semantics proposes a partial account of meaning by leaving out the its specific linguistic aspects.

3.5.3. *Frames* as the structure of knowledge

The theory of *frames* (Fillmore, 1975, 1977, 1982, 1985; Fillmore & Atkins, 1992) - together with other related notions and similar proposals - delineates a valuable and innovative proposal at the intersection between linguistics and psychology, based on the fundamental assumption that cognitive, linguistic, and experiential knowledge cannot be rigorously delimited and separated. Meanings, even when taken as mental constructions, have to be understood and analyzed on the background of the knowledge shared among the speakers. Every time, processes of conceptualization and comprehension are supported by cognitive phenomena, interactive practices, and the broad range of knowledge we acquire about the world. The trademark of frame semantics is its maximalist perspective on meaning: cognitive linguistics, and frame semantics especially, gives up on the distinction between dictionary and encyclopedic knowledge and recognizes how our (linguistic as well as non-linguistic) knowledge about the world is organized not in the form of single concepts, but rather in larger chunks of knowledge, which are then organized in a structured and comprehensive system. In Fillmore's view, *frames* are not just additional means for organizing concepts, but rather they show the way in which comprehension and communication take place. Semantic representations encompass the linguistic as well as non-linguistic knowledge that is involved in normal processes of communication and comprehension.[92]

Within this theoretical background, Fillmore's *frame semantics* can be viewed as an attempt to represent how specific linguistic knowledge is connected to the rest of our knowledge about the world. Fillmore proposes that a *frame* is a schematization of experience represented at a conceptual level and held in long-term memory. The frame relates the elements associated with a particular culturally embedded scene from human experience. Even from this sketched introduction to the notion of *frame*, it seems evident that one of Fillmore's concerns was to get over the idea that meaning can be described by checking elements out of a closed list of features. One of the problems associated with modelling knowledge solely in terms of check-lists is that people's knowledge regarding conceptual entities seems to be inherently relational (Fillmore,

[92] Fillmore (1982, 1985) distinguished between *U-Semantic* (*Semantics of Understanding*) and a *T-Semantics* (*Semantics of Truth*); the first is interested in explaining the nature and modalities of the formation and structure of the concepts within a theory of understanding, the latter's goal is to characterize the conditions under which individual utterances of a given language can be said to be true. Cognitive semantics is fundamentally a *U-Semantics*; its goal is to investigate the general conditions under which we produce and comprehend and not the conditions of truths of the utterances.

1985). "Thus a serious problem with viewing a concept as a straightforward list of features is that there is no obvious way of modelling how the relationships between the components of the list might be represented. The theory of frames represents an attempt to overcome this shortcoming" (V. Evans & Green, 2006, p. 223). *Frames* are coherent, global and systemic organization of knowledge, the structure within which the meaning of a word is defined;

> In a program designed for the teaching of English vocabulary to, say, student of English as a foreign language, we would surely be surprised to find the word *Thursday* introduced in the first lesson, *Sunday* in the fourth, and the remaining weekday names distributed randomly throughout the curriculum. Nor we would expect to find *father, mother, son, daughter, brother,* and *sister* separated from each other, or *buy, sell, pay, spend,* and *cost,* or *day, night, noon, midnight, morning, afternoon,* and *evening*. These words form groups that learners would do well to *learn together*, because in each case they are lexical representatives of some single coherent schematization of experience or knowledge. In each case, to understand what any one member of such group is about it, in a sense, to understand what they are all about. And since the knowledge which underlies the meanings of the words in each group is generally acquired all at once, it would seem natural to expect student to learn words together. What hold such word groups together is the fact of their being motivated by, founded on, and co-structured with, specific unified frameworks of knowledge, or coherence schematization of experience, for which the general word frame can be used (Fillmore, 1985, p. 223).

Similarly, what we know about the word banana includes information concerning its shape, color, smell, texture and taste; whether we like or hate bananas; perhaps information about how and where bananas are grown and harvested; details relating to funny cartoons involving banana skins; and so on. All these examples show, from different perspectives, that linguistics and non-linguistic knowledge go together in setting up how we structure our knowledge of the world. *Frames* develop on the basis of perceptual and physical abilities, cognitive mechanisms (such as perception, memory, etc.), and linguistic structures. Because of their articulation and the multiplicity of their components, they seem particularly suitable to represent the complex intersection between perception, language and cognition (cf. Diodato, 2015, p. 48). Frames provide a particular perspective, a particular envisionment of the world;

> frame theory is specifically interested in the way in which language may be used to perspectivize an underlying conceptualization of the world—it is not just that we see the world in terms of conceptual models, but those models may be verbalized in different ways. Each different way of bringing a conceptual model to expression, so to speak, adds another layer of meaning: the models themselves are meaningful ways of

> thinking about the world, but the way we express the models while talking adds perspective (Geeraerts, 2010, p. 225).

Fillmore himself notes that his frame semantic model shared significant properties with lexical (semantic) field theory (Fillmore, 1985; Fillmore & Atkins, 1992); at the same time, the main difference between the two is rooted in the idea of the autonomy versus the non-autonomy of meaning: "a major activity for lexical semanticists influenced by the field notion is that of cataloguing the kind of interitem relations that can be defined for the elements of a lexicon, and characterizing the kinds of lexical sets that are structured in terms of such relationships"; in turn, in cognitive semantic theories, "a word's meaning can be understood only with reference to a structured background of experience, beliefs, or practices, constituting a kind of conceptual prerequisite for understanding the meaning. Speakers can be said to know the meaning of the word only by first understanding the background frames that motivate the concept that the word encodes" (Fillmore & Atkins, 1992, pp. 76–77).[93]

The experiential as well as cultural and social ground of frames and the perspectivation operated by language, lead to a description on two levels: on the one hand, the description of the referential situation or events consists of an identification of the relevant elements and entities and the conceptual role they play in the situation or event. On the other hand, the purely linguistic part of the analysis indicates how certain expressions and grammatical patterns highlight aspects of the situation or event.[94] Frames can either be created by or reflected in language. Although Fillmore contrasts his view with that of field theories, many researches (Barsalou, 1992; Grandy, 1992; Lehrer, 1992) argued that one needs both frames and fields to adequately discuss conceptual and lexical organization, giving that language plays a role in the way in which we organize experience and, at the same time, it is always connected to an implicit broader frame. Hence, in *frame* theory and its developments we find an interesting account of how we organize our encyclopedic knowledge and of the specific role that linguistic organization may play within a broader conception of meaning.

[93] For a discussion of the relationship between lexical fields and frames see also Fillmore (1985).
[94] Examples of such enterprise are the analyses of the RISK frame and the COMMERCIAL TRANSACTION frame. For an in-depth discussion of the structure of these frames we refer to Croft and Cruse (2004, pp. 11–14), Evans and Greens (2006, pp. 225–228), and Geeraerts (2010, pp. 226–227). For an interesting application of the notion of *frame* see Barsalou (1992).

3.6. Partial conclusions and open questions

In the previous sections of this chapter, we have tried to summarize the historical trajectory that leads to the need of understanding meaning, as well as the structure of the lexicon and of linguistic categorization, as a dynamic process at the intersection of experience, language, and cognition. We will try to round off and to be more specific about this point, and to draw some conclusions for further reflections.

At the end of the previous chapter, we had come to the conclusion that linguistic meaning should be understood, at least at a theoretical level, as a web of pertinences that arises within the framework of a continuous intersection of experience, semantics, and pragmatics that would allow us to encompass both the linguistic systematic nature of meaning and the undeniable fact that meaning is, at the same time, an individual and idiosyncratic process. Nonetheless, such a definition of linguistic meaning opens more problems than it can actually resolve: what does it mean for language to be, at the same time, systemic and idiosyncratic, structured and fluid, shared and personal, and, ultimately, *langue* and *parole*? If language actually provides a form of categorization of our experience, what do linguistic categories depend on and what is their cognitive role? Evidently, we are far from being able to answer such questions and what seemed reasonable to do was to pick a starting point and to analyze how it has been dealt with in the history of linguistic ideas (at least in the recent history, on which we focused). The starting point we chose was that language is indeed structured and organized. We have opened this chapter with the observation that, even from a layperson perspective, words seem to be variously (but not randomly) associated with each other, i.e. that language exhibits some sort of structure and organization. Then, the relevant point was to establish what kind of organization language actually exhibits, at what level this organization takes place, and in what ways we can represent linguistic meaning as a structured phenomenon. Or, at least, these were some of the main questions (and, hence, lines of research) we have tried to master in the course of the argument: how can we describe linguistic meaning and what does it mean that language is a structured system?

We have opened the debate by focusing of the Saussurian view of the structure of the lexicon, focusing on associative and syntagmatic relationships in the lexicon as the axes on which the relational value of linguistic units is established; at the same time, we have tried to point out how the Saussurian perspective left open and partially unresolved the issue of whether these

relationships are established within language itself or if they are based and grounded in some sort of mental activity of the speakers or in the use speakers make of linguistic units. Lexical field approaches sanction, even if not without controversies, the possibility of describing linguistic meaning and its structure from a perspective internal to language and, hence, they advocate for the autonomy of the structure of linguistic meaning. Similarly, other structuralist approaches to language, as componential and mixed field-componential approaches, propose various ways in which the organization of meaning can be described from the point of view of language. Specific to componential approaches is the idea that the structure of linguistic categories can be described in accordance with a traditional view of category structure. Conversely, cognitive semantics emerged as a maximalist form of semantics, in which the structuralist heritage of isolating semantics within language itself is substituted with a contextualized approach, in which natural language semantics is seen against the background of the psychological and experiential canvas.

The maximalist orientation of cognitive semantics suggests a close alliance with psychological research: a theory that aspires to cognitive plausibility should ideally link up with what other disciplines have to say about meaning and the mind. Within this framework, linguistic meaning and linguistic structure have to be understood in the light of an encyclopedic account, according to which the meaning of words cannot be described only as a consequence of the internal structure of language, but, instead, it needs to encompass the non-linguistic aspects of meaning. In this sense, as we have seen, meaning is conceptualization, it is a cognitive phenomenon, and linguistic categories are often reduced to be the expression of conceptual categories. Similarly, the structure of the category (both conceptual and, consequently, linguistic) needs to respond to the "generalization" and "cognitive" commitments of the field and, thus, needs to consider the general mechanisms of cognition that underlie categorization and to respond to the principle of cognitive plausibility. The theory of prototypes became a fundamental source not just for the study of concepts but for the study of linguistic categories as well. The overall dominance of cognitive approaches introduced undeniable and irrevocable insights for the understanding of linguistic meaning, starting from the fact that a full account of meaning cannot leave out its psychological components, to the possibility of experimentally validating theoretical assumptions and cornerstones of semantic theories. At the current stage of research in multiple fields, it is undoubtful that accounts of meaning have to go hand in hand with the discoveries and advancements regarding the structure of our cognition as well as with neurological data. After all,

if our ultimate goal is to discuss the possibility for language to have a cognitive function, to understand and establish the cognitive status of language has to be one of the primary goals.

Why should we not end our discussion here, then, and entirely rely on cognitive psychology to disclose the nature and structure of meaning? Why even bother to revive and discuss structuralist approaches? If we look at structuralist approaches in the light of our understanding of meaning today, they certainly failed inasmuch as they explicitly ruled out the cognitive dimension and generally avoided to discuss the complexity of the relationship between language and thought. Nonetheless, theories of meaning that identify the semantic dimension with the cognitive one are equally partial, given that a bulk of linguistic phenomena remains unexplained if we reduce the semantic dimension to the cognitive one and exclude the more specific linguistic aspects. If we think of what we have called the process of "doing meaning" as a dynamic interaction between language "inside" the mind, i.e. language as a cognitive phenomenon, and language "outside" the mind, i.e. language as a socially and culturally stabilized web of pertinences, the theories we have analyzed so far mostly focused on either one the opposite poles of what, instead, should be viewed as a continuous and dynamic interaction. Structuralist account of meaning focused on the structure of language from a linguistic point of view without being concerned with what is "inside" the mind. Conversely, cognitive approaches have mostly been interested in what is "inside" the mind without being as much focused on the specific linguistics aspects of meaning and its social and cultural components. From these considerations about the strengths as well as shortcomings of the two lines of research, we derive the need to take into consideration fundamental claims of both research paradigms in moving forward with the study of the structure of the lexicon.

Allow us one last consideration; it is interesting to note that most handbooks in cognitive linguistics do not have a section titled *cognitive pragmatics*. Why is that? As a consequence of the encyclopedic view of meaning,

> cognitive semanticists reject the idea that there is a principled distinction between 'core' meaning on the one hand, and pragmatic, social or cultural meaning on the other. This means that, among other things, cognitive semanticists do not make a sharp distinction between semantic and pragmatic knowledge. Knowledge of what words mean and knowledge about how words are used are both types of 'semantic' knowledge, according to this view. […] This is not to say that the existence of pragmatic knowledge is denied. Instead, cognitive linguists claim that semantic and pragmatic knowledge cannot be clearly distinguished. As with the lexicon-grammar continuum, semantic and pragmatic knowledge can be thought of in terms of a continuum. While there may be

> qualitative distinctions at the extremes, it is often difficult in practice to draw a sharp distinction (V. Evans & Green, 2006, pp. 215–216).

According to cognitive linguists, linguistic meaning is based on usage and experience. This means multiple things: on the one side, it means that meaning is experientially grounded and rooted in experience (in the line with *embodied* theories). On the other side, it points to the importance of language use for our knowledge of a language itself. Usage-based approaches to meaning emerged from the study of grammar and of the relationship between syntax and the lexicon within cognitive linguistics:

> the experience of language is an experience of actual language use, not of words like you would find them in a dictionary or sentence patterns like you would find them in a grammar. That is why we say that Cognitive Linguistics is a usage-based model of grammar: if we take the experiential nature of grammar seriously, we will have to take the actual experience of language seriously, and that is experience of actual language use. Again, from the point of view of mainstream twentieth century linguistics, that is a fairly revolutionary approach. An existing tradition tended to impose a distinction between the level of language structure and the level of language use – in the terms of Ferdinand de Saussure (generally known as the founder of modern linguistics), between langue and parole. Generally (and specifically in the tradition of generative grammar), parole would be relatively unimportant: the structural level would be essential, the usage level epiphenomenal. In a usage-based model that considers the knowledge of language to be experientially based in actual speech, that hierarchy of values is obviously rejected (Geeraerts, 2006a, p. 6).

Introducing usage-based models of language as a fundamental assumption of cognitive linguistics has a number of consequences and, to some extent, it does not seem to be so obvious if looking at what we have said so far about cognitive linguistics.[95] There is a line of research within cognitive linguistics (and especially within cognitive grammar) that developed a fundamental link between cognitive semantics and the study of performance; more and more, cognitive linguistics conceives itself as a usage-based approach to language, of which the essential idea is the dialectic nature of the relationship between language use and the language system. Grammar does not only constitute a knowledge repository to be employed in performance, but it is itself a product of language use. The focus on language use redefines the language system is a dynamical way. One of the

[95] We will come back to usage-based theories of meaning and to their foundation in cognitive grammar in Chapter 5.

immediately evident consequences is that cognitive linguists have to invest in the analysis of real language use if they are to live up to their self-declared status.

In conclusion, we may say that our discussion of the nature of meaning as something we have to look for at the intersection of purely linguistic and purely cognitive theories led to the addition of two new pieces to our puzzle. In the first place, if linguistic meaning is, at least to some extent, cognitive and it is indeed structured, then we should investigate the cognitive organization of meaning by means of psychological and cognitive methods going beyond what we have encountered so far in the theory of prototypes and frames. Recent advances in cognitive linguistics and in psycholinguistics have made impressive steps forward in the discussion and analysis of words in the mind, but a fundamental question remains overall underspecified and partially unanswered: what do we mean when we talk about words in the mind? What is the "mental lexicon"? The second piece of the puzzle is a consequence of the observation that linguistic meaning cannot be completely ruled out and that it regains attention in usage-based and, as we will see, in distributional accounts of meaning. Corpus-based approaches allow, especially in contemporary research, to gain important insight on the nature and structure of meaning as well as on the mechanisms underlying language acquisition and comprehension, developing a new line of intersection of linguistic and cognitive approaches.

We have thus outlined the focal points of our next two chapters: first, what we come to know about the structure of language when looking at the mental lexicon; second, what we come to know about the structure of language when looking at language in use.

Chapter 4

The mental lexicon: the structure of language in the mind

4.1. A (quasi)blind journey to discover words in the mind

In the previous chapter, we have seen how the rise of cognitive sciences and of psycholinguistic approaches outlined a new framework for the study of the nature of language, of the structure of linguistic categories, and of the organization of language in the mind. It goes without saying that, so far, we have merely scratched the surface of a complex and multifaced field of research that is characterized by a diversity of theoretical, methodological, and disciplinary approaches. The aim of this chapter is to attempt an outline of some of the main theoretical orientations, methodological developments, and inter-disciplinary efforts involved in the problematic inquiry about the nature and structure of the mental lexicon. Hence, this chapter, as the one before, will be a historical and theoretical journey in the study of meaning taken, this time, mostly from a psycholinguistic and psychological perspective. It turns out that our journey will be anything but easy, since, in many ways, the study of the mental lexicon is itself a quasi-blind journey. Let us be more specific about the various ways in which the journey to discover words in the mind and our journey through the various proposals are quasi-blind journeys (see Figure 8 for a summary of some of the main issues we will be dealing with in these first introductory sections).

4.1.1. The wobbly definition of the mental lexicon

First of all, there does not seem to be a unique and commonly accepted definition of what the object of study actually is: what does it mean to study words in the mind? What is the mental lexicon?

After the discussion in the previous chapter, we now know something about the fundamental working assumptions underlying psycholinguistic and psychological studies of language both from the first and second cognitive revolution. In the first place, meaning is

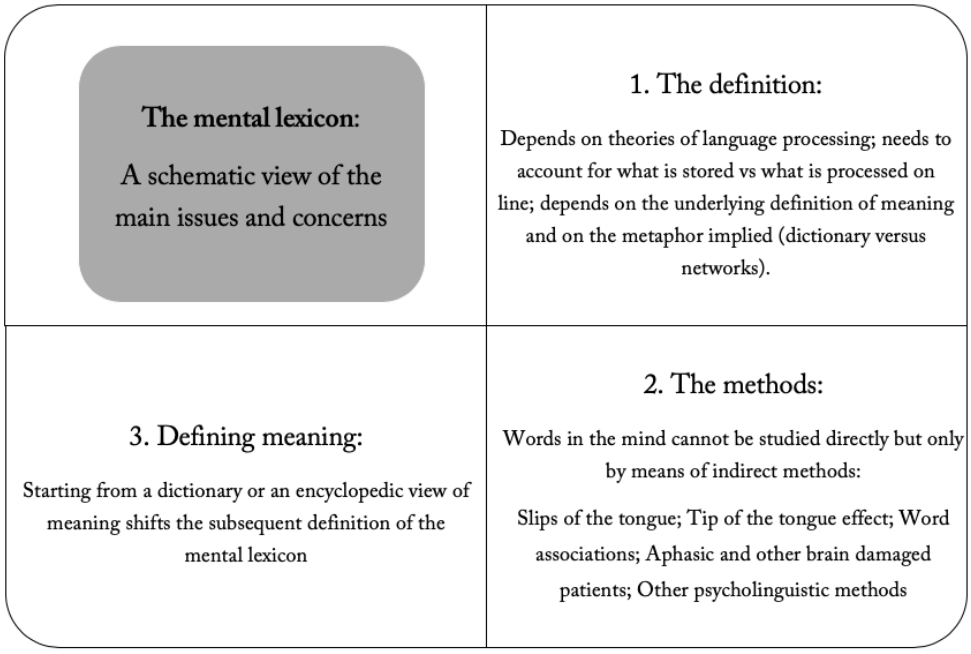


Figure 8 .A summary of the main issues and concerns in the psycholinguistic study of the mental lexicon.

conceptual: on the one side, this means that the meaning of words is a mental and psychological fact, and as such it has to be investigated; on the other side, many psychological views of meaning share the idea (even if not all with the same strength) that people translate the real world into concepts and words simply label these concepts or subset of these concepts. Hence, the meaning of words overlaps to a large extent with the concept, providing a more or less direct window on thought. In the second place, meaning is understood as a combination of various kinds of information (phonological, morphological, syntactic, semantic). As we will see in more details, the various theories we will consider focus on how these different components are organized and how they might interrelate on the basis of general (or language specific) mechanisms. In the third place, models of language knowledge, of language processing, and of linguistic interactions, need to be cognitive plausible; i.e. language and linguistic meaning need to be understood in relation to and in accordance with the general architecture and mechanisms of cognition. A forth assumption, generally implicit and underspecified, is that, for the most part, people have similar beliefs about what words mean; it is generally assumed that there is sufficient overlap between the meaning of words in the mind of different speakers (cf. Aitchison, 1987, p. 50).

Within this theoretical framework, researchers soon realized that one way to address language from a cognitive point of view was to investigate how language is processed in the mind, i.e. what are the mechanisms involved in language processing and how these processes need to be structured. Hence, the core question underlying the nature and organization of language in the mind involves the explanation of how individual users engage in everyday processes of language comprehension and production. Within psychological research paradigms, the sub-discipline of psycholinguistics developed a wide range of approaches interested in outlining, for example, models of language acquisition, language development, language processing, and language deterioration. Within this framework, models of language processing seek to discover the principles that govern how speakers match up concepts with words and retrieve those words from the lexicon, how hearers break up a string of sounds into words and find the grammatical patterns in that string, what constraints memory places on these processes, why speech errors happen, and so on. Especially at the beginning of the research enterprise, most models sought to explain the "step-by-step" processes involved in the production and comprehension of words and, especially, of utterances.

What does this have to do with the mental lexicon? The vast majority of psycholinguistic research on the mental lexicon involves the investigation of lexical processing, from which the nature and organization of lexical representations is inferred. In practice, research on the mental lexicon can be equated with the study of language activities, and the notion of the mental lexicon is a by-product of the efforts made in the description of the processing of language. In order to be able to efficiently process language, we need to have some sort of storage of linguistic (or linguistic and non-linguistic) information in our minds. The main practical reason that brought psycholinguistic researchers towards developing theories of what has been called the mental lexicon is that we seem to be extremely fast at retrieving and processing words, i.e. at "locating" them within a very large pool of possibilities.[96] Furthermore, intuitively we would say we have a mental representation of a word; people can be said to possess words, to acquire words, to use

[96] We can recognize a word in our native language in less than 200 millisecond and experiments of "speech shadowing" – in which participants wear headphones into which a stream of speech is played and are asked to repeat what they hear – show that people genuinely process the words they hear since they correct mistakes (of which they were often unaware) (Aitchison, 1987; Eric Heinz Lenneberg, Chomsky, & Marx, 1967; Marslen-Wilson & Tyler, 1980, 1981).

words, and to lose words. This implies that words, or linguistic information more broadly, need to be located somewhere in the mind and, more precisely, somewhere in long-term memory.

According to Coltheart (2001), the term "mental lexicon" was introduced into psycholinguistics in Ann Treisman's (1961) unpublished doctoral thesis. According to the author, in order to understand the nature of the monitoring carried out during selective attention tasks, it is necessary to make hypotheses regarding the nature of the word identification systems. These hypotheses involved the postulation of a mental lexicon (or, as she calls it, of a mental dictionary).

Hence, based on the principles of linguistic research described above and on the study of language processing, we have come to a first possible definition of the object of inquiry: the mental lexicon is thought of as a mental storage that contains information regarding a word's meaning, its pronunciation, morphology, syntactic characteristics, and so on, and that, at the same time, is organized for the purposes of information retrieval for speech production and comprehension.

What are the elements that we need to have stored in order to process language? For example, according to Jackendoff (2002), the lexicon is the store of the memorized elements, which are not only words but include regular affixes and stems, phrasal units and constructions. We will extensively come back to Jackendoff's position later in our discussion. As for now, we are just interested in pointing out the distinction made by the author between what is stored in the mental lexicon (in long-term memory), and what aspects of language, on the contrary, can be constructed on-line in working memory. This distinction between what needs to be stored and what can be computed on-line is at the basis of many accounts of what are the actual elements that are stored in our mental repository of linguistic knowledge. One possibility is that we store only regularities, while exceptions are computed on-line thanks to a separate set of rules. Nonetheless, the opposite view has received a lot more support: the mental lexicon is "the repository of unpredictable facts about language" (M. L. Murphy, 2010, p. 128). According to this view, what is arbitrary (in the sense that it cannot be predicted), and therefore cannot be generated, needs to be recorded in the mental storage (cf. for example M. L. Murphy, 2003, 2010). In general, the distinction between what is stored and what is processed, and hence the need for a selection of elements that end up in the mental lexicon, is typical of generative and modular views of language processing, according to which, besides the lexical elements, we also have a separate store for generative and transformational rules that operate on the lexical storage.

These few general principles aside, there is overall little agreement on what kind of information is stored and on how many levels different kinds of information might be organized and represented. For example, according to the principle of compositionality of meaning, we do not store words as whole, but rather we store their constitutional components at different levels of organization. Just to give a taste of the kind of questions and reflections we will be dealing with when discussing the information contained in this mental storage, it is worth mentioning that many approaches in psychology and psycholinguistics endorse the view that at least two different levels of organization and storage need to be present in order to account for linguistic processing: a phonological level and a separate semantic-syntactic level (cf. Caramazza, 1997 for a review and discussion).

Starting from these considerations, we can now say something more precise about the definition of the mental lexicon and about the kind of inquiry it implies: a definition of the mental lexicon arises from our knowledge of the processing of language in the mind; it should allow us to make a distinction between the stored aspects of meaning and what aspects are processed on-line in working memory; it should account for the different levels in which words can be decomposed; it has to be supported by an adequate theory of phonology, morphology, syntax, and semantic; it should put up with cognitive models of memory structure and functioning; it should provide an understanding of the organizational principles of language to account for the needs of speech production and comprehension. It seems thus clear why so few general definitions of the mental lexicon have been attempted; in general, at least to the best of our knowledge, most authors do not begin or end their papers by openly stating that the mental lexicon is such or such a thing or another (cf. Jarema & Libben, 2007b).[97] More often, in early as well as recent publications, the mental

[97] Probably one of the few straightforward and comprehensive definitions we found in the literature articulates the notion of the mental lexicon as follows: "The mental lexicon is the cognitive system that constitutes the capacity for conscious and unconscious lexical activity" (Jarema & Libben, 2007a, p. 2). This definition is in line with what argued so far, in the sense that it encompasses the two natures of the mental lexicon; "it seems relatively uncontroversial to suggest that the mental lexicon is a cognitive system. The term system suggests a degree of functional integrity that we know to be characteristic of lexical activity. Words may or may not be represented in a similar manner in the mind/brain. There is no question, however, that words are linked to one another. The term cognitive system highlights this fact, while making no claims regarding the extent to which the mental lexicon is monolithic and the extent to which it is structurally or functionally encapsulated." (Jarema & Libben, 2007a, p. 2) A fundamental component of the mental lexicon is that of being lexical activity and not of enabling it. For example, word comprehension, or saying and writing words, are lexical activities and take place is the mental lexicon. Hence, linguistic information and processes are comprised within the idea of the lexicon as a cognitive system. We will not further develop this perspective in detail, since it seems to lack of a method of investigation. Nonetheless, we will see how network

lexicon is a general umbrella-term to which researches refer to indicate the study of the structure and processing of language from a cognitive perspective. In what sense the mental lexicon is a lexicon and in what sense it is cognitive depend on the specific outline of any given theory under scrutiny. We have thus outlined a first sense in which the journey into the discovery of the mental lexicon is a blind one: its object of investigation can hardly be grasped as a unitary phenomenon and it can be theoretically and methodologically broken down into various aspects concerned both with the kind of information stored (do we store words, part of words, phonemes? What about syntactic components? Do we store collocations or entire phrases?) as well as with the type of linguistic mechanisms at task (e.g. word association, language production, language comprehension). As a consequence, the variety of possible approaches and directions of inquiry creates a maze of interrelated paths, often difficult to accurately reconstruct.

4.1.2. The metaphor of mapping the London Underground

A second way in which the journey into the mental lexicon is blind depends of the fact that, at least at the current state of research, we cannot access our object directly. We have already touched this point when we briefly mentioned that it is only by means of observing the linguistic processes that we can infer something about the nature and organization of lexical representations. Let us elaborate this point by means on an enlightening metaphor suggested by Aitchison's (1987) in her discussion of the possibility of building a model of human cognition in general and, in the specific case, of the human mental lexicon;[98]

> we are trying, then, to produce a diagram of the connections in the mental lexicon which is in some respects comparable to a plan of the London Underground. However, there

approaches to the mental lexicon will progressively shift towards a view of the mental lexicon as an activity more than as a storage.

[98] This point is of great interest to us and we will have to come back to it in the following sections of our discussion, since it will become a theoretical focal point of language simulations and Natural Language Processing paradigms of research. Since Aitchison wrote her book in 1987, although there has been a massive development of work in both theories and methods in the field of Natural Language Processing, the fundamental issues put forward by the author seem to stand even nowadays; can we find out something about the mental lexicon by making one ourselves? At the current stage of our knowledge about the human mind, it is impossible to say whether a computer simulation actually tells us how the mind works. What we can come to know is, in fact, that the models we build reproduce the human output and, on this ground, we can assess that the simulation was successful. What we do not yet know is if the mechanisms at play in the simulating models are the same mechanisms at play in the human mind. (Aitchison, 1987, pp. 28–29). We will have the chance to come back to a more detailed discussion of the assumption, methods and achievements of NLP later.

> is one way in which this mental map is quite different. We can go down into the Underground and map the connections between stations. But we cannot view the connections in the mind directly. We are instead in the situation of observers who could watch passengers entering and leaving train stations but could neither enter the system nor communicate directly with the travelers (Aitchison, 1987, p. 30).

The fundamental difference between drawing a map of the London Underground and of the mental lexicon is that we cannot directly access the mental lexicon. We know what comes out of the Tube, e.g. the linguistic spoken and written outputs, and what goes in it, e.g. the perceived linguistic inputs, but we do not know how these are processed and what kind of connections and train shifts the passengers (in our case, words) have to make in order to reach their destination. Furthermore, we do not know if the system is a homogeneous one, e.g. it is constituted only by trains, or by set of systems of different nature.

Together with the objective difficulty in observing words in the mind, what this analogy points out is that human behavior is often the end result of the interaction of quite different systems, components, and interrelated elements. The mental lexicon, in the sense of the processes involved in speech production, speech comprehension, memory retrieval, is indeed a kind of human behavior and, as such, is likely to be based and depend on multiple parallel processes.

These observations leave us with the question of what methods can be used to study the structure and organization of words in the mind. What we can mostly do in order to study the mental lexicon is to observe which passengers get on and off the trains and time them; stepping out of the metaphor, various observations of natural speech and multiple methods in psycholinguistics allow us to do this with words. For example, we can observe the connections between words in the mind by looking and the slips-of-the-tongue that occur in natural speech. In addition, much evidence has been drawn from the observation of various kinds of aphasic patients and, similarly, the tip-of-the-tongue effect seems to tell us something about the nature of the mental lexicon. As for psycholinguistic methods involved in the study of the structure of linguistic knowledge in the mind, free word association tasks provide a fundamental way of establishing the kinds of connections made between words. Measuring reaction times on priming effects is another way of opening a window on the mental lexicon.[99] These, among others, are some of the standard ways in which we can track the movements of our travelers in and out the systems.

[99] At this point of our discussion, we are merely listing the methods involved in the study of the mental lexicon. Further considerations about the advantages and limitations of each will be made explicit in the course of our discussion.

Nonetheless, given that these methods provide only an indirect observation of the systems involved, they require further interpretation in order to account for the structure of the mental lexicon as a whole. These interpretations are, for the most part, educated guesses shaped in terms of suggestive metaphors: birdcages, treasure-holds, attics, libraries are just some of the traditional ways in which the human memory (and the human memory for language with it) has been described (cf. for example Aitchison, 1987, pp. 31–32). The metaphors we use to describe the mental lexicon are, at the same time, tools and encumbrances; in fact, on the one side, they certainly are a means by which we organize in a coherent way the amount of data and observations acquired. On the other side, the kind of metaphor we use in the description often entails broader consequences and is tightly bound to some underlying fundamental assumptions about the nature of linguistic meaning.

In the study of the mental lexicon, there have been two main metaphors that are often presented as the opposite poles in the understanding of the nature and structure of words in the mind. On the one side, the mental lexicon is often described as a kind of dictionary. On the other side, the mental lexicon has often been described as a network; as we will see, networks can be of many kinds and they can support different views of what type of representations are stored. The core aspect of networks is that of the interconnectivity of the elements in a web of links with various strengths. In our discussion, we will try to organize the various proposals and studies of the mental lexicon around these two main metaphors.

4.1.3. The nature of linguistic meaning further complicates the definition of the mental lexicon

We can expand Aitchison's metaphor of the London Underground in order to make a further point:m let us pretend that we live somewhere where living beings are composed of various natures, e.g. they may be in part humans, in part aliens, in part something else. Or they may be made by just one of these possible components. We simply do not know and, at least for now, there seems to be no secure and straightforward way to distinguish between these components. Furthermore, it may be the case that different beings are required to take different trains and connections based on their composition. Thus, we are staring at these various passengers going in and coming out of the underground, but we have no clue about their nature and, even less, about what kind of trains or connections they may have taken. Hence, our guess about what kind of

journey one specific passenger may have taken is largely based on our assumption about his/her nature and about the kind of connections that that precise nature implies.

The situation with words is largely similar. The case of free words associations, as we will see, is emblematic; we know that there is a certain degree of consistency in what other words are elicited by a certain word. What we do not know, for example, is the nature and entity of the connection between the target word and the elicited words. Even more complicated is to pinpoint to specific components of the target and elicited words as means by which the connections may have been established. For example, the word *dog* is often associated with *leash, husky, beagle, cat, woof, pet*; what does it imply for our knowledge of the mental lexicon to know that these words are associated? What kind of relationships is there between *dog* and *leash*? Is it the same relationship that binds together *dog* and *cat*, or *dog* and *husky*? Is the connection a linguistic one or does it depend on other kinds of experience?

We have come to outline a third way in which our journey into the mental lexicon is blind: even if we would able to grant ourselves access to some portions of the mental lexicon, we still need to interpret the data we obtain in order to spell out the nature and type of organization underlying how we store words in the mind. Much of our understanding about the connections between words in the mind and the core assumptions underlying models of language production and comprehension depend on our theoretical understanding of the nature of linguistic meaning in general, of the internal architecture of words, and on how we theoretically cope with the relationship between language and thought. In the previous chapter, we have discussed and compared the dictionary view and the encyclopedic view of meaning. It has been argued (see for example V. Evans & Green, 2006) that the notion of the mental lexicon is intrinsically dependent on a dictionary view of meaning.[100] According to this perspective, then, dictionary knowledge relates to knowing what a word means, and this knowledge represents a specialized component, the "mental dictionary" or "mental lexicon". The common assumption is that the meanings stored in our minds can be defined, much as they appear in a dictionary. According to the dictionary view, the core meaning of a word is the information contained in the word's definition (and hence in the mental lexicon) and this is the proper domain of lexical semantics. Nonetheless, formal theories

[100] For a discussion of dictionary versus encyclopedic view of meaning see chapter 3 and references therein.

differ quite considerably on the issue of what kind of information is represented in the lexicon, such as grammatical information relating to word classes, phonological information, and so on.

With the rise of second-generation cognitive linguistics, the definition of the mental lexicon had to be re-approach in the light of a new understanding of meaning. Cognitive lexical semantics takes the position that lexical items are conceptual categories: a word represents a category of distinct yet related meanings that, for example, exhibits typicality effects or can be thought of as being organized on the base of *image-schemas* or *frames*. If the meaning of words is "fuzzy", then it may be extremely difficult to characterize the entries in a person's mental lexicon understood as a dictionary. We will see how the properties of the mental lexicon will shift in accordance with the theory of meaning the underlies it. For example, we need to know whether it is possible to assign a firm definition to a word or whether words inevitable have fuzzy meanings. The answer to this question will obviously affect our view of how people represent words in their minds. For example, as Aitchison remarks, taken that word meaning does not have a straightforward and clear definition as would be the case in traditional-formalist account,

> people sometimes have to interweave their own observations with information presented by others, since biological taxonomies and cultural beliefs may clash with instincts. For example, children often find it hard to believe that a spider is not an insect, a whale is not a fish and a bat is not a bird. And in Papua New Guinea, the Karam people of the upper Kaironk valley do not regard the cassowary as a bird, even though to us it obviously is one. It is unclear how such 'facts' get integrated into a person's overall view of a word's meaning (Aitchison, 1987, p. 60).

Moreover, in some cases it seems also very difficult to separate the meaning of a particular word from the whole situation in which it occurs. The frames (or scenarios) provide a background into which one fits the details of the present situation. From the point of view of the mental lexicon, how do we store such background information? Is it a different kind of storage compared to the one of strictly linguistic information? In the case of a positive answer to this last question, how do linguistic meaning and background knowledge communicate? As it might already be clear, the encyclopedic view of meaning - which rethinks the nature of linguistic categories and opens to a much more informative storage that comprises both linguistic and non-linguistic meaning – calls for a substantial reshaping of the properties and structure of the mental lexicon.

4.1.4. Brief summary and outline of the chapter

So far, we have mostly outlined some of most problematic aspects in moving towards the study of the mental lexicon. However, these issues are not limited to the notion of mental lexicon; as Whitaker (2006) points out reporting a quote from Ebbinghaus (1885, trad. English 1913) in the preface to the inaugural issue of *The Mental Lexicon*:

> Two fundamental difficulties arise in the way of the application of the so-called Natural Science Method to the examination of psychical processes: (1) The constant flux and caprice of mental events do not admit of the establishment of stable experimental conditions. (2) Psychical processes offer no means for measurement or enumeration (Ebbinghaus, 1913 , Section 9, Original Version 1885).

In addition, the study of words in the mind poses a further complication: in order to understand the nature and organization of language, researches have to be conducted at the intersection of, at least, linguistics and psychology. Specific issues regarding the study of the mental lexicon include: 1) the difficulty in establishing a specific definition of the mental lexicon that goes beyond very general assertions about it being a mental storage of linguistic information. This is mostly due to the fact that the definition of the mental lexicon is not a direct one, but rather a by-product of observations and theories on how we process language; 2) the overall impossibility of having a direct access to the mental lexicon and, hence, the need to infer properties of this hypothesized mental storage from a variety of differently interpretable data and experiments; 3) the fact that the definition of the mental lexicon and the specific aspects of language processes considered for investigation vary accordingly to the theoretical stands on the nature of linguistic meaning.

In the remaining of this chapter, we present how various theories and accounts contributed in shedding light onto this blind journey. The various sections of this chapter will try to address some of the main contributions for each of the points addressed so far and, hence, they will be organized as follows (see Table 1): in the first place, we will start by addressing the notion of the mental lexicon as a dictionary. This will allow us to discuss models of the mental lexicon sprouting from generative and modular paradigms (with a focus on phonological and morphological studies) (*3.2.*). In the second place, we will analyze models of the mental lexicon based on the metaphor of semantic networks (or cob-web approaches), generally considered to be better predisposed to account for an encyclopedic view of meaning; we will predominantly focus on the concepts of

<table>
<tr><th colspan="5">The study of the mental lexicon as a by-product of research on language processing</th></tr>
<tr><th>Metaphor</th><th colspan="2">Language processing</th><th>Theoretical assumptions</th><th>View on meaning</th><th>(Main) Methods and data</th></tr>
<tr><td rowspan="2">The mental lexicon as a dictionary (3.2.)</td><td colspan="2">Serial processing (3.2.3.)</td><td rowspan="2">Generative linguistics
Modular view of the mind
Symbolic processing (information theory)
Compositionally of meaning (3.3.1.-3.3.2.)</td><td rowspan="2">Dictionary view of meaning</td><td rowspan="2">Slips-of-the-tongue
Tip-of-the-tongue effect
Aphasic patients
Phoneme monitoring</td></tr>
<tr><td colspan="2">Interactive processing (3.2.4.)</td></tr>
<tr><td rowspan="3">The mental lexicon as a network (3.3.)</td><td rowspan="2">Networks of local representations (3.3.2)</td><td>Quillian's model (3.3.2.)</td><td rowspan="2">Cognitive linguistics (second revolution)
Computational neurosciences
The science of networks (3.3.1.)</td><td rowspan="3">Encyclopedic view of meaning</td><td rowspan="3">Free association data (3.4.1.)
Reaction times of semantic priming
Computer simulations
Network analysis</td></tr>
<tr><td>Large-scale networks (3.4.2.)</td></tr>
<tr><td>Neural-networks (3.3.1 - 3.3.3.)</td><td>Networks of distributed representations (3.3.3.)
Words as cues-to meaning (do we actually need a mental lexicon?) (3.3.4.)</td><td>(same of local representations plus)
Connectionism
Parallel distributed processing (3.3.1. -3.3.3.)</td></tr>
</table>

Table 1. Summary of some of the main approaches in the study of the mental lexicon

parallel processing, spreading activation, and, in part, to the more recent developments in the study of semantic networks within the framework of network studies (*3.3.*). By means of operating this distinction between dictionary views and network views, and by means of connecting the first with dictionary and atom-globular views of meaning, and the latter with encyclopedic view of meaning, we are establishing a set of oppositions that, in reality, is definitely not so clear-cut and that is not always there in the studies we will analyze. Nonetheless, setting it as a guiding principle of organization of our discussion would allow us to try to encompass and summarize a broad range of studies within a scaffolded theoretical perspective. In the last section, we will discuss the free words association method, its advantages and pitfalls, and its most recent applications. Out of all the methods used in psycholinguistics, we selected to address free word association tasks because they are still a largely used method, although the theoretical discussion about what kind of data they provide remains highly underspecified (*3.4.*). An intermediate section will discuss the question of whether we actually need anything like a mental lexicon at all to account for language processing, especially in the light of recent advances in cognitive psychology and in Natural Language Processing (NLP) (*3.3.4*).

4.2. The mental lexicon as a dictionary

The term *mental lexicon* was originally conceived as an artifact of a pervasive metaphor: that of a dictionary in our heads. Given that the mental lexicon is, in very general terms, a sort of storage of information about words that allows us to speak and comprehend, what does it mean for the mental lexicon to be "as a dictionary"? Do we look up words in our mind by checking them in alphabetic order? Ones we locate a word in the mental lexicon, do we find its definition as we would in a printed dictionary?

What these questions actually address are the following issues regarding the organization and nature of the lexical storage: to argue that the mental lexicon is "as a dictionary" implies that it has to be organized according to a certain organizational principle (whether it would be phonological, in stacks or piles, maybe as a thesaurus, and so on) and that we should be able to say what kind of information is actually stored in it (phonological, morphological, semantic, syntactic, and so on). Thus, on the one side, thinking of the mental lexicon as a dictionary means to search for the principle by means of which the lexicon is organized. For example, when we look up the word *color*, are we more likely to find it connected with *colostomy, colostrum, colotomy, colt, column*, or with *colors, colored, colorable, colorful, colorist*, or maybe with *blue, orange, pink, glow, hue, intensity*? It is hard to imagine that a strict alphabetic order could account for the organization of the words in our mind or that it would be of any actual help in the retrieval of words for both comprehension and production. Mistakes like slips of the tongue suggest that even if the mental lexicon turns out to be partially organized in terms of initial sounds, the order will certainly not be straightforwardly alphabetical. Other aspects of the word's sound structure, such as its ending, its stress pattern and the stressed vowel, are all likely to play a role in the arrangement of words in the mind. Furthermore, not all instances of slips of the tongue can be accounted for by means of their phonological aspects; many examples of slips of the tongue involve substitutions, modification, and errors in the (semantic) content of the exchanged words (Aitchison, 1987; Fay & Cutler, 1977). Less straightforward is the issue of whether morphological derivations and inflections need to be present in the lexicon and whether or not they could be a possible principle of organization. Similar problems arise when we consider the syntactic-grammatical components of a word and its semantic component (taking semantic as the concept the word refers to). How do we store the information that the word *color* can be both a noun and a verb and that, therefore, it

will follow different rules given the sentential context in which it occurs? Or, maybe, do we store *color*, *colors* (as a plural noun and as a verb in the third singular person), and *colored* as independent, maybe related, units in the mental lexicon?

Within the framework of the mental lexicon as a dictionary, the separation between what needs to be stored in the mind and what can be processed online has been the focus of much of the generative enterprise interested in the topic and it is fundamentally related to the idea that the set of grammatical and compositional rules can be separated from the semantic (meaning) component of language. Are morphological inflections and derivations (like *colored, colorable*, and so on) stored in our mind or do we create them on-line while speaking and comprehending? Likewise, should we account for fixed collocations and phrases, like *color-blind*, *color-coded*, as stored in our mind? What about fixed phrases that alter the meaning of the single words involved, like *grey area, green thumb*, *out of the blue*? If the information stored in the mind for the word *grey* would be that of a dictionary, such as that it is an adjective denoting "a color intermediate between black and white, as of ashes or an overcast sky" (Oxford Dictionary Online)[101], it would be very hard to account for all the usages we encounter and we can make of it. We know a lot more about the meaning of words than what can be stored in any possible printed dictionary: we know with which other words a word can or cannot be used; we know how words relate to each other; we know the syntactic pattern of a word; we know that the same word can be pronounced in many ways although remaining the same word. But if we limit ourselves to a dictionary view of meaning, according to which only certain elements of word meanings need to be stored, we can outline very general consideration. Early models of the mental lexicon arose in continuation (and later in contrast to) the generative and the modular paradigms.[102] Within this theoretical framework, some very general

[101] https://www.lexico.com/en/definition/grey

[102] In the study of the mental lexicon, we can spot a detachment from the original Chomskian framework similar to that observed within the rise of cognitive semantics (cf. paragraph 3.6). A substantial departure from the original Chosmkian framework is the study of language processing itself as carried out by psycholinguistic approaches. Discussing what aspects of an utterance must be stored in long-term memory, and what aspects can be constructed on-line is a question of performance, rather that of competence. Considering language in terms of the cognitive processes that grant its understanding and production can be already seen as a shift towards a theory of language performance that could lead to even major reorganizations of the theory of grammar (Jackendoff, 2002). Furthermore, a linguistic theory of syntax, even in the standard conception of a theory of syntax as a formal device that generates well-formed strings in a language, is not itself, and was not intended as, a theory of how language is actually processed in cognitive terms. Even less direct is the possible cognitive interpretations of a strict generative transformational grammar (Marslen-Wilson & Tyler, 1981). From a psycholinguistic perspective, the cognitive implications of linguistic theories of syntax, and of linguistic theories in general, for language processing are an empirical question that has to be evaluated in respect to the data on how the processes actually function.

intuitions about the structure of the mental lexicon can be identified. First and foremost, for language comprehension and production to take place, the mental lexicon needs to be organized into subsystems.

4.2.1. The mental lexicon is organized into subsystems: *lemmas* and *lexemes*, production and comprehension

The fundamental assumption underlying the search for an organizational principle of the mental lexicon often implies the possibility of separating different aspects of linguistic meaning and, hence, to isolate one as the foundation of the organization of the lexicon. Research stemming from generative and modular backgrounds tend to hypothesize different levels of representation in the form of separate subcomponents containing phonological, orthographic and semantic information about words (Allport & Funnell, 1981; Emmorey & Fromkin, 1988; Forster, 1976; Morton, 1982).[103] The majority of the models that we will take into consideration agrees on distinguishing between, at least, two levels of representation: the level of the *lemma*, which is the level of syntactically and semantically specified lexical representations, and the level of the *lexeme*, with carries information about phonological and orthographic representations (originally set forward by Kempen & Huijbers, 1983). But, as we will see, the distinction among various other levels of representation can go even further.

Nonetheless, if on the one side we can theorize multiple independent levels, on the other side, in producing and comprehending sentences, we are constantly faced with the interrelation of sound, meaning, and syntax. Hence, if we assume the existence of two (or more) independent components that allow us to process specific linguistic information, we need to set out a framework

[103] Also, in approaches not directly aligned with a dictionary view of the mental lexicon, such that of Aitchison, we can find direct claims of the separability between phonological aspects and semantic aspects of the lexicon: "words, can be viewed as coins, with meaning on one side and sound on the other. The human sound-meaning coin, the word, is more sophisticated than that of most animals because of one crucial factor: humans have an apparently innate realization that the coins are not just indivisible whales. Instead, they can be analysed" (Aitchison, 1987, p. 203). The sounds of words seem to be easily unhooked from their meaning and word class. According to Aitchison's view, the two parts are inherently detachable because the component dealing with the semantics and word class in arranged conveniently for production, whereas the phonology is organized primarily for speedy recognition. The mental lexicon seems to be a mixed system which has found a workable compromise between the requirements of production and those of comprehension. The component that is required first in each case has imposed its demands on the organization. Production begins with the semantics and syntax, so these are arranged to suit production. Recognition begins with phonology, so this is organized to suit recognition.

that accounts for their interaction or interrelation. We will see how the various models will mainly divide into serial versus interactive models.

The features of discreteness and of seriality of the levels of representation and of the processes involved seem to be a clear heritage of the modular architecture of the mind, combined with the general view of language as computational information processing, as theorized for the purpose of developing large and efficient computer programs. As Marr (1976) explicitly states,

> Any large computation should be split up and implemented as a collection of small sub-parts that are as nearly independent of one another as the overall task allows. If a process is not designed in this way, any small change in one place will have consequences in many other places. This means that the process as a whole becomes extremely difficult to debug or to improve, whether by a human designer or in the course of natural evolution, because a small change to improve one part has to be accompanied by many simultaneous compensating changing elsewhere (Marr, 1976, p. 485).

Thus, in order to account for cognitive processes of large complexity, as the processing of language, we need to work within a modular framework. The system of language processing is organized into discrete subsystem; on the one side, we can distinguish between a subsystem that accounts for speech production and a subsystem that accounts for language comprehension. On the other side, as we have seen, we need to distinguish between various discrete levels of representation that organize the subsystems of production and comprehension into further subsystems that deal with phonology, morphology, syntax and semantics.

In addition to the computational framework set forth by the birth and development of computer engineering and of artificial intelligence, the close interrelation of biology and psychology contributed in directing the framework of the study of language processing and of the mental lexicon to a strong modular perspective. Much evidence in support of this view came from patients affected by damages to the brain who seemed to maintain some modules intact, while others not. Lexical knowledge about a word is therefore not stored in "one" place, but is represented in different sub-lexicons (Emmorey & Fromkin, 1988). In that there is a good correspondence between the elements of psychological model and human's anatomy, then one might suppose that anatomical facts could inform the psychological theory. Towards the end of the nineteenth century a number of researchers tried to construct information processing models of brain functions, largely using data from patients with brain damage (Lichtheim, 1885; Wernicke, 1874). The fundamental assumption that guided the study of various forms of aphasia

in brain damaged patients was that localized damages would give rise to behavioral observations that diverged from the performances of normal language users. A vast amount of data from aphasic patients seems to speak for functionally localized areas of the brain that, in turn, would inform us on the organization of cognitive subsystems (for a discussion cf. among others Morton, 1981; Shallice, 1979). Neuropsychological case studies to brain damaged patients largely supported the hypothesis of the modular architecture of the mind and of the language processes: patients with localized damages show significant selective impairments of specific processes (such as the translation between different language codes) or levels of representation (such as selective phonological impairments), suggesting a discrete modular view as the most fruitful way of accounting for the empirical data, in that it allows to uphold the functional separability of different components of language, or of other information-processing systems (Allport & Funnell, 1981; Caramazza, 1997).

Other methods in the study of the mental lexicon seemed to further support a modular view of language processing. The primary source of psycholinguistic data in language production has come from speech errors or slips-of-the-tongue, which are taken to reveal the properties of the language system and of how the human word-store works. In the first place, there are multiple kinds of speech errors that can be roughly categorized into selection errors and assemblage errors, of which the first are the most relevant to characterize the mental lexicon. Among selection errors, we can distinguish between errors based on meaning, on sound, and on sound and meaning (cf. Aitchison, 1987). Such evidences have been employed to support the lemma/lexeme distinction and sound-based errors especially have been used to argue for a phonological organization of the mental dictionary. By means of working on malapropisms, phonological slips-of-the-tongue, Fay and Cutler (1977) argued that the basic arrangement of the lexicon is by phonological segments for the purpose of aiding comprehension (for a discussion cf. Cutler & Fay, 1982; Hurford, 1981).

Further support for the lemma/lexeme distinction and for the possibility that language may be phonologically organized for the purposes of words' retrieval comes from the tip-of-the-tongue (TOT) phenomenon. The TOT phenomenon is a state in which one cannot quite recall a familiar word but can recall words of similar form and meaning: we fail to retrieve a lexeme (the phonological/orthographic component) in the context of successful retrieval of its lemma (the syntactic and semantic component). It has been demonstrated that while in the TOT state, and before recall occurred, participants to experiments had knowledge of some of the letters in the

missing word, the number of syllables in it, and the location of the primary stress (R. Brown & McNeill, 1966) and, hence, that the TOT state can be viewed as a window on the processes involved in word retrieval: responses which are phonologically related to the target - the word that participants in the study had to guess after its definition was provided - are more likely to be followed by correct retrieval than responses which are only semantically related. The responses most consistently associated with eventual success are those in which the phonological resemblance incorporates a root morpheme of the target, suggesting, once more, a phonological organization of the mental lexicon for the purposes of word retrieval (Kohn et al., 1987).

Another useful example of the distinction between lemmas and lexemes comes from studying in language production. Garrett (1975) proposes that sentence production, which he defined as the translation of an intended message into a linguistic form, was accomplished in a number of serially ordered independent stages. In Garrett's model, syntactic processing takes place in two stages, a "functional" level stage during which lexical representations and their underlying grammatical relations are constructed, and a "positional" stage during which a representation consisting of phonologically specified morphemes is constructed. The two-stage assumption explains a number of regularities in speech errors, including the fact that word exchanges and sound exchanges have different characteristics. Word exchanges typically occur with words of the same syntactic class; they can occur across a span of several words; and the participating words do not need to be phonologically similar. In contrast, sound exchanges can involve sounds from words that differ in syntactic class; the participating sounds are usually close to one another; and they are usually phonologically similar. Garrett accounts for these differences by assuming that word exchanges take place at the functional level and sound exchanges at the positional level. Garrett's proposal, among others, highlights the need to distinguishing between different modality-dependent lexical codes stored in the mind and accessed sequentially.

Given these first insights, what can we preliminary conclude about the structure of the mental lexicon? The modular assumption about the functioning of the mind led to some fundamental implications for the structure of the information stored in it. We have seen how, within this theoretical framework, in order to account for various kinds of psycholinguistic data - researches on brain damaged patients, observations of speech errors, and of induced as well as spontaneous tip-of-the-tongue states - researchers have often found a comfortable ground in the idea that the lexicon is stored in the mind by means of a subset of lexical systems that are discrete,

autonomous, and modality-dependent. Hence, a first main feature of the mental lexicon is that it is compositional:[104] words are decomposed in the mind, and each component is an independent lexical code (phonological, orthographic, morphological, syntactic, semantic), that is separated from the other lexical codes as well as from other cognitive codes and that accounts for various aspects of language processing, both in production and in comprehension. Words are not stored "as wholes"; on the contrary, their various components are stored in separate lexical storages from which we retrieve and combine (on-line) the elements we need to understand an utterance or to produce it.

As we have seen, the main distinction between the components stored is the mind in the one between lexeme and lemma; the first in concerned with the sensory properties of words, i.e. the phonological and the orthographical, whereas the latter should comprehend syntactic, maybe morphological, and semantic components. How these two main levels of distinction, as well as their sub-levels, are connected for the purposes of communicating depends on the specific model of production or comprehension, with possibilities ranging from a totally independent and serial view of language processing, to models that allow for the interconnectivity of the various modules.

A second feature of the definition of the mental lexicon as a dictionary is that it needs to be organized in order to facilitate access to it. This means two main things: on the one side, we need to establish the order in which the subsystems are accessed based on the observable

[104] Another important insight comes from compositional semantic approaches applied to the notion of mental lexicon as a dictionary: if words are built up from a common pool of "meaning atoms", then words related in the mental lexicon have atoms in common. But researches about the mental lexicon have moved far beyond the point of the basic arguments underlying the idea of a componential description of meaning; although the theoretical distinction between componential and holistic accounts remain fundamental in the framework of a cognitive theory of semantics, it is less relevant when we turn to consider the cognitive processes and mechanisms underlying word recognition and language parsing. Models of language comprehension and production differ on whether they endorse a componential or a non-componential account of meaning, but they are in agreement of the fact that semantic, syntactic and phonological/orthographic forms constitute independent levels of representation and that these levels of representations are probably accessed sequentially in the course of language processing (Caramazza, 1997). A further step is to discuss whether, within a given representational level, say the phonological one, the phonological representation of a word is stored as a whole or as a combination of phonological trait, that in turn can be either processes on-line to constitute the word, or already stored off-line as they are combined into phonological words. An important point has to be made regarding the hunt for semantic primitives in the mind and their possible effects on language processing. Many researchers have tried to check if word comprehension could give some clues on the existence of basic meaning components, which, as we have argued, cannot be observed directly. The general assumption is that, if words that are stored in disassembled state and recomposed on-line for the purposes of speaking, then words which are likely to have a large number of components should take longer for the mind to process than words that have fewer. Experiments designed to evaluate potential cognitive effects of basic meaning features do not seem to support an atomic globule viewpoint of the organization of the mental lexicon (Cutler, 1983; J. D. Fodor, Fodor, & Garrett, 1975; Kintsch, 1974).

characteristics of the language processes. For example, the majority of the speech production models, which address how we go from conceptual meaning to spoken or written linguistic output, argue for a two-stage process, in which firstly lemma's features are accessed and, only later, they are connected to the appropriate lexeme representations. On the other side, we should be able to establish how each level is organized and how this organization influences the one of the other levels; in lexical production, it would make sense for the lexicon to be organized so to facilitate the access to semantically (or syntactically) related words, no matter their phonological representation. Following this line of thought, the opposite would happen with speech comprehension, in which the first elements to be processed are the sensory stimuli. As we will see, not all models endorse a view of separated lexicons for production and comprehension but, instead, hypothesize the existence of just one mental lexicon of which the different processes operate.[105]

Let us briefly recap: if we want to understand how we comprehend and produce language, we need to investigate how linguistic information is stored in long-term memory, its format and organization, and how it can be accessed for the purposes of word recognition and speech parsing. Similarly, we need to identify what aspect of language can instead be processed on-line, taking the load off of the mechanisms of storage and retrieval. Hence, we finally came to the point of asking: how is language processed? What kind of mental storage do we need in order for a given process to take place?

[105] In the literature about language processing and the mental lexicon, it has been largely argued, on the basis of the dissociations between receptive and productive abilities, for the existence of independent lexical representations for 'inputs' and for 'outputs' (Morton, 1979). Such a position requires the linguistic information to be stored in the mind twice, as it needs to be organized for the purposes of production and as it is best accessed for comprehension. Most of the data provided in support of this view are also compatible with the view that there is a single language system (composed of various lexical codes) on which the processes for comprehension and production operate (Allport & Funnell, 1981). Otherwise differently specified when discussing the single contributions, we will mostly be concerned with models that assume the existence of one lexical storage (made up by the different codes) on which two different kinds of processes operate, namely for producing and comprehending speech.

4.2.2. Serial models of language processing and the structure of the sub-lexicons

How do we manage to go from a heard utterance to the conceptual understanding of it, and, vice versa, from the concepts we want to express to the uttered sentence? At an intuitive level, in understanding a string of language we start with an auditory stimulus (or a visual one) that we need to parse in order to get to the next step of analysis, where we can access the syntactic and semantic components of what we initially heard (for a schematic of two-stage models see Figure 9). The distinguishing feature of a serial model of language comprehension is the assumption that the flow of information between the different components of a processing system is in one direction only, from the bottom up (for a discussion cf. Marslen-Wilson & Tyler, 1980). Similarly, for the purposes of speech production the flow of information is strictly top down: a first stage is concerned with conceptual preparation, to which follows the retrieval in long-term memory of the grammatical encodings for the selected conceptual information, and, at last, phonemes, metrical frames, and stress patterns are retrieved from the mental storage in order to produce the final phonological encodings (for a discussion cf. Schriefers & Vigliocco, 2001). In both cases, a processor at any one level has access only to the stored knowledge which itself contains and to the output from the immediately preceding stage. Thus, for example, a word-recognition component can base its recognition decisions only on the knowledge it contains about that level of linguistic representation: context-effects in word-recognition could not be located in the word-recognition

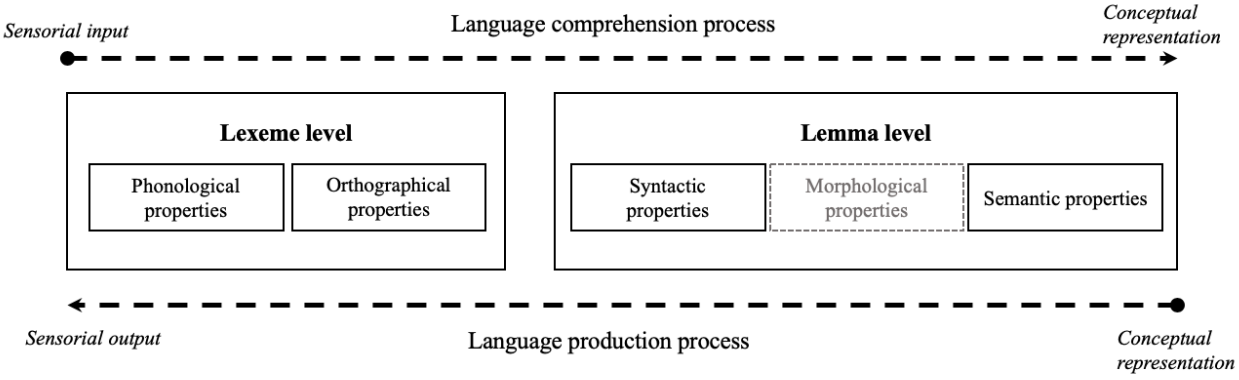


Figure 9. Simplified two-stage serial model of language production and comprehension and schematic representation of the (main) modules at each level of representation. (Morphological properties are not always considered within two-stage models and, hence, are represented in a shaded color).

component, but at some later stage, in which the appropriate constraints could be used to select among the several words that might be compatible with a noisy sequence of acoustic-phonetic input.

These processing assumptions have a profound influence on the ways in which a model implements the further assumptions that are made about the types of mental knowledge that need to the distinguished. First of all, serial models of language processing require each distinct knowledge type to be instantiated in a separate processing component, functioning as a computationally independent processing level. This has long been the position assumed in the analysis of the various components of the mental lexicons. The majority of the work has focused on the lexeme level of representation of the mental lexicon, targeting the phonological (Fay & Cutler, 1977; Forster, 1976; Napps & Fowler, 1987) and the orthographic storages as well as the need to account for grapheme-to-phoneme conversion (Allport & Funnell, 1981; Coltheart, 2004; Emmorey & Fromkin, 1988; Fromkin, 1985).

We have already said something about the phonetical level of representation in the previous section; for example, according to Fay and Cutler's (1977) perspective on the mental dictionary, words are arranged by phonemic structure, in a left-to-right manner, and are based on a distinctive features system. Those words that begin with the same phoneme are listed together, those that have the same second phoneme form a subcategory of that class, and so on:

> words […] are looked up in the dictionary on the basis of this phonological "spelling", using a left-to-right convention. So, for example, to locate cat (/kaet/), the device goes first to the section reserved for one-syllable words beginning with /k/, then to the subsection of words beginning /kae/, and then finally to words beginning /kaet/. It is here that the device will find an entry for cat with the attached information about syntactic category and meaning (Fay & Cutler, 1977, p. 513).

The major partitioning seems to be by number of syllables, with stress pattern as a second categorization within syllable categories.

In addition to the phonological representation, each lexical entry must include its orthographic representation and a set of rules for grapheme-to-phoneme conversion which permit reading aloud and writing to dictation. That phonological and orthographic representations are distinct parts of the lexicon is supported by the fact that they can be differently affected by brain damages and, especially, by the large amount of data that has been collected from deep dyslexic

patients, i.e. formerly skilled readers who appear to have lost the ability to translate between graphic and phonological codes (for a review on the relevant neuropsychological studies cf. Allport & Funnell, 1981; Emmorey & Fromkin, 1988). Models that include a separate component of the lexicon containing orthographic representations differed on whether or not there are connections between the phonological and orthographic representations of words. For example, the *logogen model* (Morton, 1979) rejects any direct connection between these two systems: some phonological representations must be contained in the auditory logogen (a logogen is a word detection unit that receives stimulus information from the analysis systems and semantic and syntactic information from the cognitive system) and orthographic representations must be coded in the visual logogen, keeping the two levels of representation strictly separated. Conversely, in Forster's (1976) model of the lexicon, all the information about a word, stored in separated sub-lexicons, is collected in a single master file. Each entry in each sub-lexicon includes the addresses which connect it to the other various representations, i.e. a word's phonological representation includes its orthographic, semantic, syntactic addresses.

A further level of storage in the lexicon in that of the morphemes (Garrett, 1980; Jackendoff, 2002; Sandra, 1994). For example, Garrett (1980) shows that affixes and word stems exhibit different patterns in speech errors, suggesting that there appear to be different constrains regarding their processing. At a more theoretical level, Jackendoff (2002), starting from the premise that lexical items stored in the lexicon may be bigger or smaller than grammatical words, suggests a distinction between the storage of productive and semiproductive morphology. On the one side, the outputs of productive lexical rules, which are totally regular except when blocked by irregular forms, are likely to be constructed on-line in working memory starting from stems and affixes stored in the lexicon. On the other side, outputs of semiproductive rules, which are only partially regular, as the irregular forms of the English past tense, must be listed in long-term memory as we encounter them, since they cannot be fully predicted given the lexical rule.[106]

[106] The idea that, together with lexical representations, we also store lexical and grammatical rules has not been addressed directly in our discussion and received overall less attention in the literature about language processing. This may be due, at least to some extent, to the methodological difficulties we would encounter if we tried to describe in a variable, or to measure with reaction times, a certain lexical or grammatical rule. Furthermore, as we have previously mentioned, the psycholinguistic study of language processing set soon aside the possibility of a direct link to the generative grammar and its set of transformational rules. Nonetheless, we think that an example of the need to account for both the linguistic items as well as the storage of linguistic processing rules can be found in Aitchison's (1987) proposal of a lexical tool-kit containing instructions and guidelines as to how to make up new words, and in

Dictionary models of the mental lexicon face substantial difficulties in accounting for a wide range of both linguistic and cognitive phenomena that seem to be part of our everyday use of language, such as the fact that we can infer the meaning of new words we do not previously know or the possibility of creating new words. For instance, this is revealed by the hesitations and struggles in the discussion about whether and how we store morphological and semantic information in the mental lexicon. When it comes to semantic representations, serial models of language essentially maintain that the access to the level of the lexeme is an autonomous process, operating solely on the basis of sensory information without inputs from higher order components in the language processing system. In the process of speech production, lexical access to a content word consists of two independent and serially ordered stages. In the first - semantically driven - stage, the so-called lemmas are retrieved, lexical items that are specified with respect to syntactic and semantic properties, but not with respect to phonological characteristics. In the second stage, the corresponding wordforms, so called lexemes, are retrieved. This implies that access to a content word involves an early stage of exclusively semantic activation and a later stage of exclusively phonological activation.

Similarly to researches in speech comprehension, in the analysis of speech production much effort has been devoted to the articulation of the structure and content of lexical-phonological representations (Garrett, 1980; Levelt, 1989; Levelt et al., 1991), whereas less specified was the content and the processing structure of lemma-level representations (Bock & Levelt, 1994; Caramazza, 1997; Jescheniak & Levelt, 1994; Roelofs, 1992, 1993). The clearest proposal concerning the processing structure of lemma representations is that of Roelofs (1992), which was subsequently adopted by Bock and Levelt (1994) and Jescheniak and Levelt (1994). The model distinguished among three levels of representation: the conceptual level, the lemma, and the lexeme. The conceptual level represents lexical concepts as unitary nodes in a conceptual network. The meaning of a word is given by a set of labeled connections between a concept node and other nodes in the network. Each lexical node is connected to a lemma node; in this model, lemma nodes are modality independent units that are connected to a set of syntactic nodes specifying such properties as grammatical class, gender, auxiliary type. Each lemma node is connected to a lexeme node which, through its connections to segmental nodes, specifies the phonological (and

order to do this it makes use of back-up information attached to each word, showing how it can be split up, and other words to which it is related.

orthographic) form of a word. While still providing a strict serial account of language production and maintaining the modular separation of the lexical components, this model shifts from the original dictionary view of the mental lexicon, in which words are stored and searched by looking them up into a list organized according to the principles operating at a certain level of representation, towards a network approach, in which the principle of spreading activation will become the lead searching-engine in the mental lexicon. We will become more familiar with these notions in the next few sections of the chapter.

In conclusion, within the theoretical framework analyzed so far, the lexicon includes strictly linguistic semantic representations, separated from non-linguistic knowledge, to which no access is granted by the representations of lower level components, hence divesting syntax and semantic, and contextual aspects of language with them, from any role in decoding or encoding conceptual information into linguistic units. As we have tried to show, the assumptions of autonomy require the definition of the characteristic processing units in terms of which each component conducts its analyses. The effective validity of the local structure of the language processing depends on the definition of level-by-level processing units, for it is the properties of these units that determine the points in the analysis of the input at which one processing component will be able to transmit information to the next, and, equally important, the points in which this further component will itself be able to start analyzing the input it is receiving, and itself generate an output to subsequent levels of the system.

These claims about the temporal ordering of sentence processing operations are overall too strong, since they require that semantic interpretation be delayed until the syntactic clause or sentence boundary had been reached. But semantically based disambiguating or facilitatory effects on lexical or syntactic processes can be observed well before the point reserved to them in classical serial accounts (Marslen-Wilson & Tyler, 1980). Even if we leave momentarily aside all the issues concerned with establishing which parts of the conceptually available information are going to be encoded, and with defining the mechanisms that convert the conceptual information into a format that is suitable for the linguistic formulation process, a long-standing issue in language production is how to characterize the mapping between conceptual and grammatical encoding, respectively the first and the second stage of the process. First is the question of whether conceptual preparation takes language specific properties into account, i.e. if the structure of language is involved in the selection of the conceptual material. Languages differ in which conceptual and formal properties

need to be realized as a detail of the sentential form (a common example is whether the gender needs to be expresses, as in Italian, or does not, as in English). Second is the question of whether conceptual information permeates the processes occurring during grammatical encoding beyond providing its input (Schriefers & Vigliocco, 2001). In the next session, we will see how, already within a modularist framework, a completely discrete and serial model is hardly justifiable and does not allow to account for a conspicuous amount of empirical data that shows the interaction between the various levels of linguistic representation.

4.2.3. Interactive models of language processing: from dictionaries to networks

Interactive models of language processing allow for the sort of interference between the levels that is observed in everyday language processing, and they provide the theoretical framework for evidences of a genuine interaction between sentential context variables and the process of spoken word recognition. Interactive models do not place *a priori* limitations on the ways in which different knowledge sources can communicate during language processing. Instead, they assume a more flexibly structured processing system, such that analyses developed at any one level of knowledge source can, in principle, made be available to affect the operations at any other level of knowledge source. In fact, this seems to be what actually happens in normal language understanding: not only we are able to recognize a word far before it has been completely uttered, but we seem to develop an interpretation of an utterance as we hear it and we use this information to influence the recognition of subsequent words in the utterance. Evidence provided by speech shadowing techniques - in which participants are asked to immediately repeat what they hear - shows that listeners can develop at least a preliminary syntactic and semantic analysis of the speech input before the actual completion of the input (Marslen-Wilson, 1973, 1975, 1985).

At an anecdotical level, the experience of simultaneous translations seems to point in the same direction: if syntactic and semantic comprehension would be possible only at the end of a strictly serial process, it would be hard to justify how simultaneous translation could be possible at all, given that the comprehension of a sentence occurs a lot earlier than the sentence has been completely uttered. In a trivial way, we know where a sentence is going, and we are able to infer, with a good level of accuracy, which words will follow the one we are listening or reading now. This type of observations leads to the claim that the local structure of the processing system permits

syntactic, semantic, and contextual interaction with word-recognition decisions (being word-recognition the level of phonological/orthographic detection and decoding). By the time a word had been identified, there should be no asymmetry in the availability of semantic as opposed to phonological information about the word, and, therefore, no necessity that semantic attribute-matching decisions be delayed relative to phonological attribute-matching. The message-level interpretation of the input is initiated right from the beginning of the utterance and directly affects all aspects of the system's operation. In more technical terms, syntactic, semantic, discourse-based processing analyses interact with word-recognition processes upon which the detection of words depends.

Admitting an effect of context sentence implies that even the first words of an utterance must be evaluated with respect to their discourse context, which provides the framework for the on-line analysis of the input. According to the *cohort model* (Marslen-Wilson & Tyler, 1980, 1981), word-recognition is an interactive system that works as follows: the individual recognition elements are assumed to be directly accessed from the bottom-up during speech recognition. Early in a word, even from the first phoneme, a group of recognition elements will become active, which corresponds in size to the number of words in the language that begin with the sound sequence that has been heard up to that point. This preliminary group of word-candidates is referred to as the *word-initial cohort*. The subsequent time-course of word recognition will depend on the ways in which the word-initial cohort can be reduced in size until only a single word-candidate is left. On the one side, this is done by means of the acoustic-phonetic information that becomes available in time; as more of the signal is heard, it will tend to diverge from the initial specifications of more and more of the members of the initial cohort. On the other side, these recognition elements interact with contextual constrains. The semantic and syntactic properties of the members of the word-initial cohort need to be assessed against the requirements of the sentential context in which the word to be recognized is occurring. If this can be achieved, then those recognition elements whose properties do not match the requirements of context can be

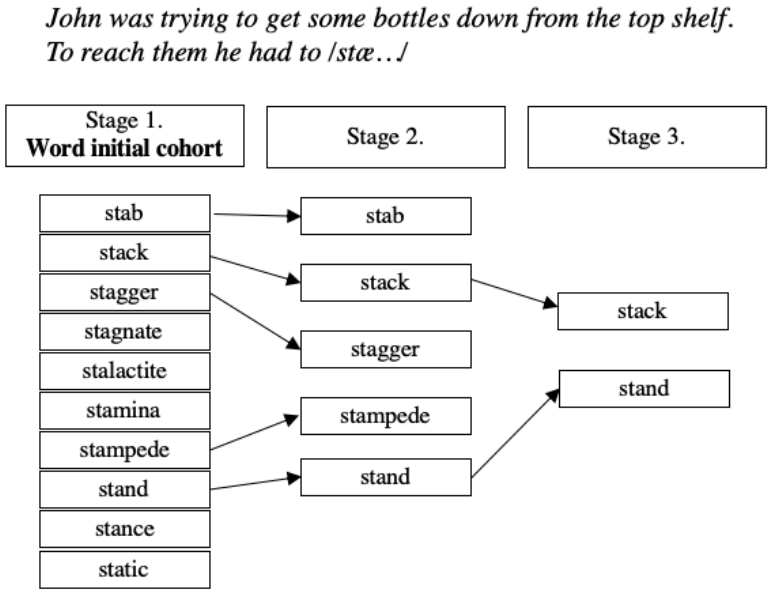


Figure 10. Representation of Marslen-Wilson's example of a word initial cohort and the progressive stages in the word-recognition process. (Marslen-Wilson, 1981; cf. Aitchison, 1987)

dropped from the word-initial cohort. An example provided by the authors is the following: if we hear the sentence "John was trying to get some bottles down from the top shelf. To reach them he had to /*stæ…*/" our word-recognition processor will estimate the word-initial cohort, i.e. the set of words that begin with the sound sequence /*stæ*/, like *stag, stance, static, stamp, statute, stack, stand*, *standard, stagger, stab,* and so on. It is clear that if sensory information only is used to make the word-recognition decision, then we word *stand*, which is the actual target, could not be identified until all of it had been heard (see Figure 10 for a representation). If syntactic constrains are used in the decision process, then the set of word that are nouns could be immediately excluded, since they could not occur at that point in the utterance. Finally, if interpretative constrains also contribute to on-line word recognition, then the word could be identified even earlier, given their plausibility in the discourse context. After these restrictions, only two candidates remain, *stand* and *stack*, which are the members of the cohort for /*stæ*/ that both match the sensory input and fully satisfy contextual constrains. Thus, as soon as the listener can determine that the next sound is nasal, he or she can select the word stand (Marslen-Wilson & Tyler, 1981, p. 111). This kind of on-line integration of sensory input and of the utterance and discourse context requirements is necessary to explain the speed and earliness of spoken-word recognition, which seems to happen

within 300 milliseconds or less from the first heard sound of the word. The process of speech recognition is essentially data driven, i.e. the initial set of possible analyses of the input is determined from the bottom up, and especially from the phonological aspects of the sensory input. This word-initial cohort delimits the decision space within which further analyses, incorporating contextual factors[107], can take place, allowing for a continuous interaction between top-down and bottom-up information.

In the process of language production, "the process by which a semantic representation of a sentence-to-be-spoken is translated into a phonetic representation that can guide the articulatory musculature and produce speech" (Dell, 1986, p. 238), the translation of meaning to sounds involves a selection of words from the semantic mental dictionary that are then ordered in adherence with the rules of the grammar of the speaker's language (syntactic encoding) and that must be specified in terms of their constituent morphemes (morphological encoding), that, in turn, must be spelled out in terms of their sounds (phonological encoding). The process of language encoding outlined by Dell (Dell, 1986, 1990; Dell & Reich, 1981) is useful to our discussion in that it binds together the need for maintaining separate levels of representations with the interconnectivity assumption of network approaches and the retrieval mechanism based on *spreading activation*. Network approaches will be the focus of the next section. Here, we will only briefly present the main characteristics of the model in order to draw some general conclusion, while postponing a more in-depth discussion of the mechanisms and assumptions of network approaches to the next section. It seems to us, that Dell's model, among others, represents an interesting middle ground between strictly modular views and network approaches. The model is based on the idea of a network that contains nodes at three levels: lemmas, lexemes, and phonological segments. The lemma node represents the lexical item as a syntactic/semantic entry, and it is directly connected to the conceptual structure (which is deposited in an even separate level) and to syntactic information. Below that, the lexeme node is a single unit representing the phonological form of the word. The lexeme node is connected to phonological segments. Each

[107] The results of studies conducted within the cohort model approach show that the semantic dimensions of processing are dominant throughout a sentence, and that they have significant effect on monitoring responses even at the very beginning of the test-sentence, where the syntactic dimension has only marginal effect. This means that those semantic aspects of processing that interact with word-recognition decisions are essentially fully developed at the beginning of the sentence. Semantic constrains operate in word-recognition by enabling word-candidates to be rejected because they map inappropriately onto the meaning representation that is available at that point in the utterance (Marslen-Wilson & Tyler, 1980, p. 41).

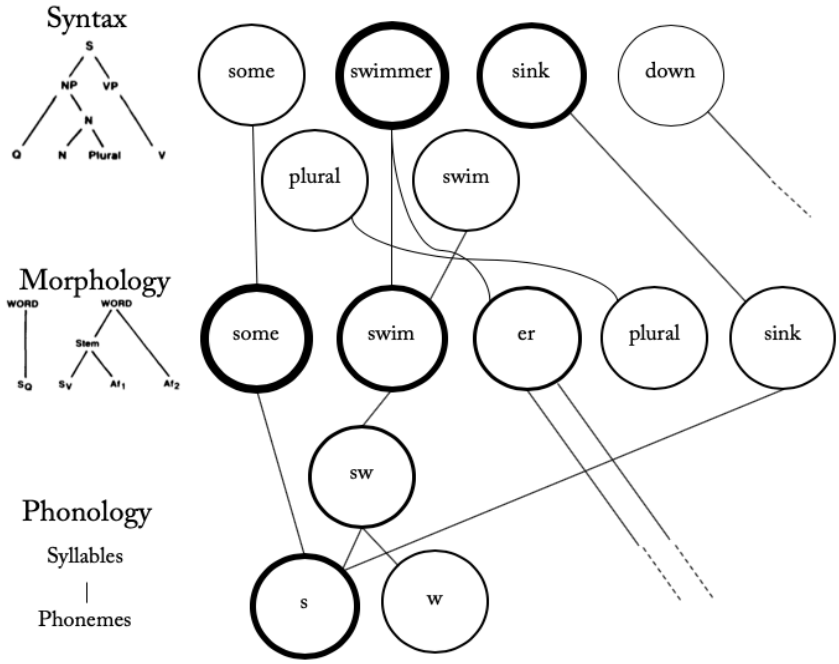


Figure 11. A simplified version of a portion of Dell's network representing the production of the sentence "Some swimmers sink". The thickness of the node represents the level of activation (the thicker the line is the more active the node) and the dashed-ending lines represent those paths that are not continued in the subsequent level. (Adapted from Dell, 1986, Figure 1, p. 290)

level stores nonproductive knowledge and is associated to a set of generative rules that define the combinatorial possibility of the units at each level. The network structure assumes that, at each level, each particular unit is only represented once. So, for example, at the phonetical level, the sound /æ/ is a single unit and every word that contains this sound connects to this unit. All connections are assumed to be bidirectional, i.e. they allow activation to spread both upwards and downwards. The retrieval of the form of a word involves the activation of the lemma, followed by a spreading activation process that ultimately activates phonological segments[108] (in Figure 11 we show a simplified version of Dell's representation of the structure of the levels and of the network).

[108] We briefly report how the spreading activation process works in the framework of this model: "The defining components of spreading activation are spreading, summation, and decay. When a node has an activation level greater than zero, it sends some proportion of its activation level to all nodes connected to it (spreading). This proportion is not necessarily the same for each connection. When the activation that is sent out reaches its destination node, it adds to that node's current activation level (summation). For the sake of simplicity I assume no thresholds, saturation points, or other nonlinearities in the spreading process. It is necessary, however, to include a passive decay of activation over time to keep levels down. Specifically, activation is assumed to decay exponentially toward zero. These operations, spreading, summation, and decay, apply to all of the nodes in the lexical network at all times, regardless of whether the node is part of a representation (tagged or not)" (Dell, 1986, p. 287).

What is relevant to our discussion at this point is that, even within a modular and generative framework, an interactive view of language processing must be considered in order to account for many aspects of language production and comprehension.

Hence, one of the fundamental innovations that took place already within the standard view of the mental lexicon as a dictionary and, contemporarily, in connection to the new frontier of cognitive sciences, was the development of connectionist and network approaches, which firmly pointed out that, in order to understand how language production and comprehension work and in order to establish the stored components of words, we need to account not just for the structure of the levels, but rather mostly for the interaction between them. As we have seen, processes that seem straightforward, as comprehending a word, are far more complex that we might have expected. In normal speech, it is highly unlikely, if not just impossible, that we can hear each phoneme: sounds are often altered by their neighbors in quite radical manners, and speech sounds are often mixed to other noises in the outside world. And even if we figure out how we are able to split up the stream of speech into words, we are still faced with the problem of identifying what these words mean, a task very difficult to account for in serial models of speech comprehension. In fact, we are often faced with ambiguous sentences, sequences of sounds that correspond to more than one meaning, new words the meaning of which needs to be interpreted on the spot, words that are recognized before they occur, and words that are recognized one or two words after they have been spoken. Hence, what we need to develop in order to understand the structure of language processing is a model that can account for the continuous update and flow of information between the various levels of representation and a mental lexicon which cannot be a fixed selection of stable and defined elements. This is where the notion of the mental lexicon as a network of relationships between stored items comes handy.

4.3. The mental lexicon as a network

4.3.1. A new theoretical framework: from "boxes-and-arrows" to "nodes-and-weights"

The majority of the models stemming from the generative and modular framework can be described as "boxes-and-arrows" models (Christiansen & Chater, 2001c), in which language processing is described in terms of flow diagrams: each "box" contains a set of specific linguistic information and the arrows between the boxes describe the direction in which the input and output information for each module flows. These "boxes" build up the mental lexicon, i.e. the set of specific linguistic information stored in the mind in various discrete modules containing specific kinds of information. This view is tightly linked to the definition of language and of linguistic meaning typical of generative grammar. Furthermore, it was highly influenced by the refutation of semantics as a proper field of linguistic research and the focus on the formal aspects of language. Lastly, if we look at the models from the perspective of the underlying cognitive principles, linguistic knowledge rests on innately specified, intuitive, domain-specific features of the mind.

As we have seen in the previous chapter when presenting the rise and development of cognitive semantics, advancements in cognitive research have progressively moved away from many of the principles imposed by the generative framework. Cognitive linguists from the second cognitive revolution hypothesize that language is the product of general cognitive abilities and argue that the solutions to enduring problems in semantics and grammar require abandoning the theoretical framework of the early cognitive descriptions of language. While the proponents of this shift have emphasized different linguistic problems, such as the non-independence of grammar or the non-compositionality of semantic meaning, they concur in rejecting two major tenets of the Chomskian linguistics: the separateness and specialness of language, i.e. of a specific innate mental organ for language, and the modularity of different types of linguistic information (C. L. Harris, 1992). The refutation of these fundamental architectural principles goes hand-in-hand with a new understanding of the nature of linguistic meaning: neither the form nor the meaning of expressions can be adequately described without the reference to speakers' encyclopedic knowledge, their construction of mental models, their ability to map concepts from concrete to abstract domains, their use of super-positional representations, and the schemes to integrate multiple sources of information. If we want to understand how language is processed in the mind,

language cannot be viewed as independent from the general principles of cognition, from general-purposes learning mechanism, from the neuro-anatomical organization of the brain, and from the rest of our knowledge about the world.

These shifts in cognitive paradigms, which emphasized the central role of the interrelation and interconnection of the various aspects of language and the centrality of semantics (understood as general encyclopedic knowledge) for a cognitive account of language, brought to a substantial change in how language processing had to be understood and how it could be modelled.[109]

The initial wave of cognitive sciences tended to see theories of human information processing as borrowing from theories of machine information processing. "Symbolic" processing on general purpose digital computers has been the most successful method in designing practical computer. It is therefore not surprising that cognitive sciences, including the study of language processing, has aimed at modelling the mind as a "symbolic" processor. Traditional "symbolic" models of information processing and mainstream computer sciences are based on the intellectual tradition that views human thought as a matter of symbolic manipulation, e.g. communication is the manipulation of logical statements expressed using variables and operations and language processing consists of a sequence of serial, rule-governed steps (for a critical review cf. Christiansen & Chater, 2001b; O'Reilly & Munakata, 2000). It is easy to note how "boxes-and-arrows" models and the consequent view of the mental lexicon develop the same line of thought that underlies the general claim about the nature of thought as a manipulation of symbols; symbolic manipulations occur in a serial order and we have a more or less clear idea of how information

[109] The question of the legitimacy, or maybe the need, of computational models of the mind and of its functioning is a long-standing issue in the history of the cognitive sciences. In order to understand how the brain thinks, researchers make use of many types of methods, first of all brain imaging. At the same time, computational models based on biological properties of the brain can provide and important tool for understanding the complexity of the human brain and of human behavior. For example, such models can capture the flow of information from our eyes reading, up to the parts of the brain activated by the different word meanings, resulting in an integrated comprehension of the text. Of course, we do not know much about what is actually going on, at least not yet. Computational models seem to allow us to explore the underlying mechanisms of the hypothesized structure, which can be implemented on a computer, manipulated, and tested. An extremely active field of research that specialized in computationally modelling cognition is now known as computational cognitive neuroscience; researches in the field simulate human cognition using biologically based networks of neuronlike units. One fundamental assumption of computational cognitive neurosciences, when compared to other movements in artificial intelligence, cognitive psychology, and cognitive sciences, is the idea that it is not sufficient to identify computational and cognitive properties, but rather that suppositions about how the mind works need to be grounded on and consistent with the biological mechanisms of cognition. The aim of many of the approaches that we will take into consideration in the following sections is to build interactive models that simultaneously include both the constrains of the working from biological facts up to cognition and working from cognition down to biological facts (O'Reilly & Munakata, 2000, pp. 1–8).

should be stored so that systems can access and manipulate it. Nevertheless, as we have seen in the previous section, "box-and-arrows" models of language processing fail to account for many psychological as well as biological data on the functioning of the brain, on how information might be actually processed, and on the structure of linguistic information in the mind. Many of the problems faced by modular and serial models of language have already been pointed out at the end of the last section; the one general problem that is worth recalling at this point of our discussion is that serial models have a hard time dealing with the interconnectivity of the modules and with data showing that various kinds of information are present at almost every stage of language encoding and decoding. Furthermore, boxes-and-arrows models are not fully explicit computational models, in that they did not actually allow to develop full blown simulations and implementations. Indeed, not only they performed very poorly in explaining and predicting people's behavior, but they also seemed not to take into consideration the neuroscientific evidence that challenges the modular assumption on the structure of the mind.

What is the alternative, then? In many fields of study, from cognitive psychology to computational neurosciences, one of the most influential metaphors for studying the structure and organization of the mind and for modelling cognitive processes was that of the network or, in Aitchison's (1987) words, of the cobweb. There is considerable diversity among network models of cognition, but all models are built up on the same basic components: simple processing elements and weighted connections among these elements. The basic elements are generally called "nodes" and the connection between nodes have "weights". Information flows between the nodes that are connected in the network; a node receives inputs from other nodes to which it is connected. In some networks the connections are unidirectional; in other cases, connections can be bidirectional. It is in the weights of the connection between nodes that knowledge is progressively build up in the network. This picture offered a new metaphor for the study of language: given that words cannot be assembled out of a common store of semantic primitives, that the "boxes-and-arrows" framework does not allow us to model the processing of language, and that the dictionary view of the mental lexicon as a closed list of various kinds of linguistic information supported by a separate set of combinatorial rules does not allow us to explain how words can actually be understood and retrieved, cognitive sciences and psycholinguistics needed a different framework in which they could describe the structure of the stored information and how cognitive processes operate on them.

There are many ways in which the general metaphor of the network has been used and implemented and it would go far beyond the interests of this chapter to discuss all of them. We will thus focus on to two main directions of research which are historically connected but took substantially different theoretical paths, especially for what concerns the form of the representations within the network. We will propose a distinction between cobweb approaches and neural network approaches, i.e. between network approaches that endorse a local and categorical representations (that we will call cobweb according to Aitchison's proposal) and neural network approaches, which propose a distributed view of representations. In very general terms, cobweb supporters claim that words are regarded as wholes which have various characteristics and enter into relationships with other words. The main mechanism of word retrieval is condensed in the notion of *spreading activation*. On the contrary, neural networks models stem from a connectionist approach to cognition and endorse a parallel distributed view of representations. One of the fundamental mechanisms involved in the growth and structure of neural networks is, as we will see, the principle of backpropagation.

The issue of how words are represented in the mind is obviously central to understanding language and it remains so for cobweb and connectionist (or parallel distributed) approaches; the repository of word-level representations is referred to as the lexicon, and many traditional approaches (including some cobweb approaches) have assumed that there is a centralized, canonical lexicon in the brain where each word is uniquely represented. In contrast, connectionist models of language processing suggest that word-level representations should be distributed across a number of different pathways specialized for processing different aspects of words. The idea of a *distributed lexicon* has been championed by those who modeled language from the neural network perspective (e.g. Plaut, 1997; Seidenberg & McClelland, 1989). We will devote the next two sections at trying to make explicit the conception of the mental lexicon that underlies the two mentioned approaches and we will see how, compared to the previously analyzed models, network models of the mental lexicon have a direct interest and application in the study of the semantic component of linguistic meaning. One of the leading questions in the literature interested in network approaches is what kind of models, or what parameters of a given model, allow us to understand (hence, simulate) how we perform semantic tasks and constitute a reasonable description of the general organization of human memory for such material (e.g. Quillian, 1967; Rogers & McClelland, 2004). At the end of our discussion, we will briefly take into consideration

how connectionist and distributed approaches to meaning can lead towards rethinking the very necessity of a mental lexicon.

4.3.2. Local representations and spreading activation

How does a network need to look like to account for human linguistic, and specifically semantic, behavior? What is semantic behavior in the first place? The literature we are about to explore takes semantic meaning as the whole of the knowledge we have about something in the world, in accordance with non-autonomous, encyclopedic views of meaning. Hence, semantic tasks are those that require a person to produce or verify semantic information about an object, a depiction of an object, or a set of objects indicated verbally. Semantic information in semantic tasks is the information that has not previously been associated with a particular stimulus object itself. Semantic tasks requiring the retrieval of semantic information are, for example, verifying that the object shown in the picture is a cat, or that the pictured object can purr, and also verifying sentences like *cats have fur*, as long as the information we are testing for is not directly associated with the particular picture or with previous sentences (Rogers & McClelland, 2004, pp. 2–3). The fundamental aspect of a semantic task is that it needs to involve retrieval of information stored in long-term memory. Hence, modelling semantic behavior has much to do with being able to come up with a proper and efficient representation of the structure of semantic memory.

Is modelling semantic processing the same as modelling semantic representations? Following McRae and Jones (2013), we can argue that, at a very general level of analysis, semantic processing models are those that focus on how learned semantic structure is used in a given task, whereas semantic representation models are those that specify a mechanism by which semantic memory is formed from experience. The two things are obviously interconnected: how we learn new semantic information depends on the current state of semantic memory, and semantic structure and process influence each other when explaining behavioral data (Johns & Jones, 2010). How a network learns semantic information will be one of the main interests of connectionist approaches to language and will be the focus of the next section. The models we will present here are mostly concerned with semantic performance.

Semantic task performance is usually thought to depend on a mediating process of categorization. This means that we need to assume that representations exist in memory, in this

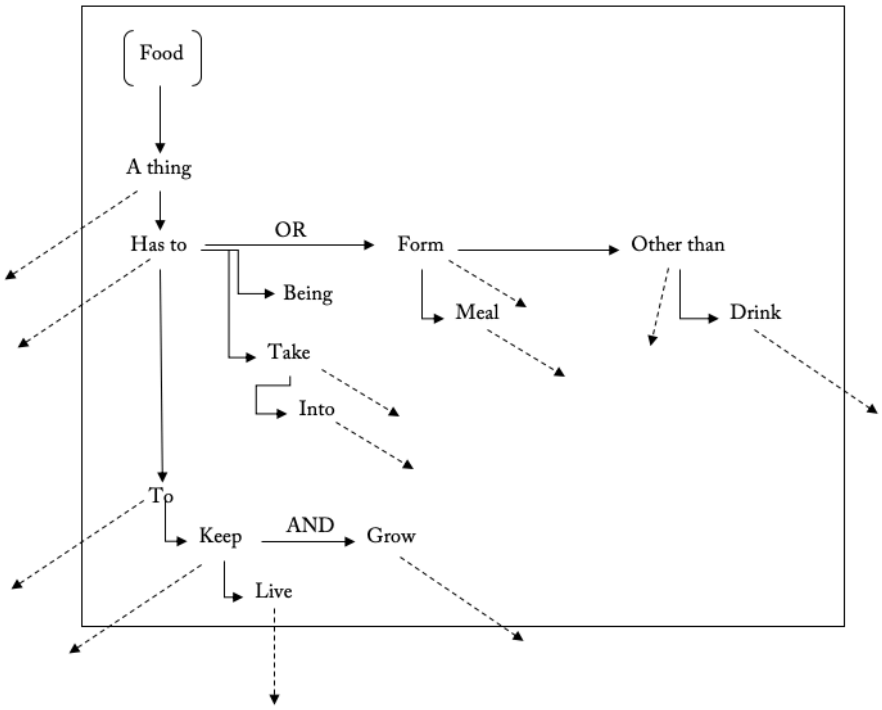


Figure 12. A simplified version of the network description of the word "food" starting from its definition: "That which living being has to take in to keep it living and for growth. Things forming meals, especially other than drink". According to Quillian's view, the connections established in the representation of each word create a plane. Arrows that leave the plane indicate connections with other planes, i.e. with the network representation of other words (Adapted from Quillian, 1976, Figure 1b, p. 415).

case as nodes in a semantic network, corresponding to each of many concepts or categories, and that information about these concepts is either stored in the representation itself or is accessible from it. Categorization-based models (which are a kind of local-representation models) of semantic performance argue that performance on semantic tasks depends on access to the relevant category representation. We are back to the question we started from: how is semantic knowledge organized?

One influential view has been that semantic knowledge is efficiently represented in stored category representations, linked together in a taxonomically organized processing hierarchy, a way that seems to be very efficient for both storing and retrieval of semantic information (Collins & Loftus, 1975; Collins & Quillian, 1969; Quillian, 1967, 1969). A key aspect of Quillian's (1967) proposal flows from the observation that category membership at each level entails a number of

properties shared by all the members of the more specific included category. The goal of Quillian's proposal was "to develop a theory of the structure of human long-term memory, and to embody this theory in a computer model such that the machine can utilize it to perform complex, memory-dependent tasks" (Quillian, 1967, p. 410). The model looked grossly as follows (see Figure 12 for a practical example of the representation of a word): the model consisted of a mass of nodes interconnected by different kinds of associative links (HAS and IS A links, e.g. a *dog* IS A *animal*, a *dog* HAS *tail*). The nodes - the relevant units of human conceptual store - are not themselves words - even if we can address them by means of using words (i.e. we can name the nodes by means of words) - but rather they are closer to "properties". The meaning of words is given by the set of nodes (hence of properties) and the relation between the nodes. There are two types of links between the nodes: the first is a direct link - the associative link leads directly into a configuration of other nodes that represents the meaning of its name word. This kind of link identifies a *type* node. The second kind of link refers indirectly to a word concept, by having a special kind of associative link that point to that concept's type node.[110] This is a *token* node. This distinction between *type* and *token* nodes is a consequence of how words are defined in ordinary dictionaries:

> For defining one word, the dictionary builder always utilizes tokens of other words. However, it is not sufficient for the reader of such a dictionary to consider the meaning of the defined word to be simply an unordered aggregation of the other word concepts used in its definition. The particular configuration of these word concepts is crucial; it both modifies the meanings of the individual word concepts that make up its parts, and, with them, creates a new gestalt which represents the meaning of the word being defined. […] In short, token nodes make it possible for a word's meaning both to be built up from other word meanings as ingredients, and at the same time to modify and recombine these ingredients into a new configuration (Quillian, 1967, p. 411).

Hence, the full representation of the content of a word consists of:

> of all the information one would have if he looked up what will be called the "patriarch" word in a dictionary, then looked up every word in each of its definitions, then looked up every word found in each of these, and so on, continually branching outward until every word he could reach by this process had been looked up once. However, since a

[110] In more details, the links need to be complicated enough to represent any relation between two concepts. Quillian identifies five different kinds of links: superordinate links (is-a) and subordinate links, modifier links, disjunctive sets of links, conjunctive set of links, and a residual class of links, which allows the specification of any relationship where the relationship itself is a concept (for a discussion cf. Collins & Loftus, 1975).

> word meaning includes structure as well as ingredients, one must think of the person doing the looking up as also keeping account of all the relationships in which each word he encountered had been placed by all earlier definitions (Quillian, 1967, p. 413).

According to this model, the definition of a word, like *machine*, encompasses everything we know about the concept "machine", starting off with the more compelling facts about machines and proceeding down to less and less inclusive facts. The order in which a certain concept-word is brought forth is from general, inclusive facts to less and less closely related ones, forming something that may be viewed as a hierarchically ordered body of information.

One first aspect of this network model, when compared to previously discussed models, is clear: the semantic information stored in the model goes past beyond that envisaged by a dictionary view of meaning, and it is by means of the different types of associative links, which connect the nodes together into configurations, that the model is able to represent the variety and richness of the content expressed in any natural language, allowing, at the same time, for an economic representation of redundancy. Furthermore, a specific feature of such model is that information is encoded in local representations: "given a network of simple computing elements and some entities to be represented, the most straightforward scheme is to use one computing element for each entity" (Hinton, McClelland, & Rumelhart, 1986, p. 77).[111] In the case of this model, each unit corresponds to a certain "property" of the word-concept.

Such an organization of memory seems to be useful in explaining how we perform in semantic tasks such as in verifying that a certain word has a certain property. Collins and Quillian (1969) suggested that one could test the idea that the mental lexicon is organized in a hierarchical way by asking subjects to verify a sentence such as *A canary is a canary* or *A canary is a bird* or

[111] The point about local representations made by Hinton, McClelland and Rumelhart in the book that sanctioned the rise of connectionist theories moves from a technical observation concerned with the way in which information is encoded in the hardware: in simulating models, such as Quillian's model, the physical network mirrors the structure of the knowledge it contains. A concept corresponds to each computing element, i.e. to each unit in the network: there is a one-to-one correspondence between concepts and hardware units. In order to avoid possible misunderstanding, it seems worth pointing out that the framework outlined by network approaches to language processing developed within a computational framework is concerned both with the technical aspects of creating such models, which involve both software and hardware issues, as well as with how these models allow to explore the microstructure of cognition. These two aspects go hand-in-hand and open the discussion to some fundamental issues concerned with the development of AI, which cannot be addressed here. For the case we are making here, i.e. we are concerned with how these models of language processing allow us to represent and describe the structure of the mental lexicon, the adherence of these models to human cognitive processes has to be discussed for each model we are taking into consideration (for a discussion of the proper treatment and appeal of connectionist approaches cf. O'Reilly & Munakata, 2000; Rumelhart, McClelland, & Group, 1986; Smolensky, 1988).

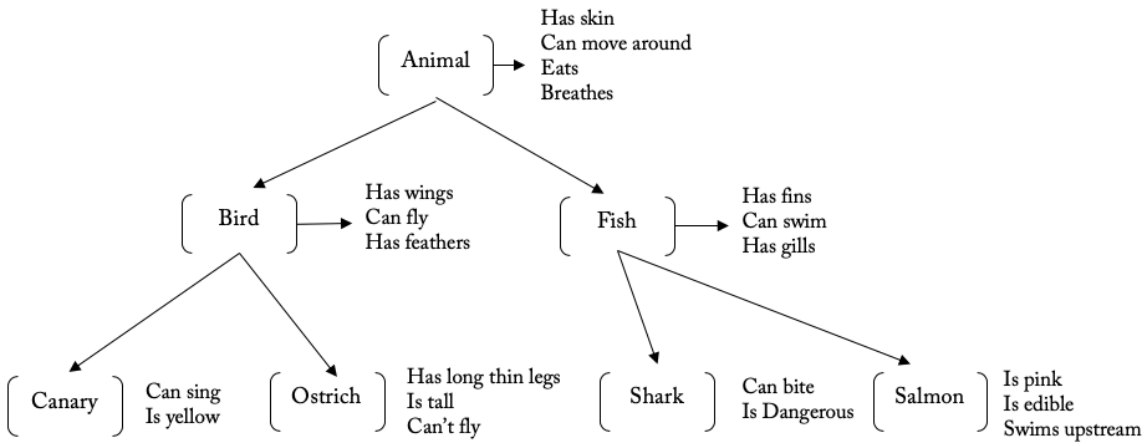


Figure 13. A representation of the memory structure of a three levels hierarchy (Adapted from Collins and Loftus, 1969, Figure 1, p. 241).

A canary is an animal, and time how long it took people to verify each sentence. In Quillian's model, each word has stored with it a configuration of pointers to other properties stored in memory (for a representation of Collins and Quillian's model see Figure 13);

> if what is stored with canary is "a yellow bird that can sing" then there is a pointer to bird, which is the category name or superset of canary, and pointers to two properties, that a canary is yellow and that it can sing. Information true of birds in general (such as that they can fly, and that they have wings and feathers) need not be stored with the memory node for each separate kind of bird. Instead, the fact that a canary can fly can be inferred by retrieving that a canary is a bird and that birds can fly. Since an ostrich cannot fly, we assume this information is stored as a property with the node for ostrich, just as is done in a dictionary, to preclude the inference that an ostrich can fly (Collins & Quillian, 1969, p. 240).

The hypothesis was that, if a person has to navigate a hierarchically organized tree of nodes in order to retrieve a semantic information, the further the person has to travel down the tree, the longer it would take he or she to verify a sentence like the ones before. These predictions turned out to be correct: subjects took longer to verify statements about general properties than about more specific properties, putatively because such properties were stored higher in the tree and thus required a greater number of inferences.

However, subsequent work has shown that reaction time in property verification tasks may be influenced by a variety of factors other than the property's putative position in the hierarchy

(for a review of some major experimental work on the hierarchical structure of the mental lexicon cf. Aitchison, 1987, pp. 80–82; Rogers & McClelland, 2004, pp. 8–13).

A further interesting point of Quillian's proposal is the intuition regarding the mechanism by means of which we consult and navigate the stored taxonomic relations. The author proposed a spreading activation mechanism that permitted the activation of a category representation to spread to taxonomically superordinate concepts. For example, when an object is categorized as a *dog*, the activation of the concept "dog" spreads to the related concept "animal", and the properties stored in the related concept are attributed to the object. The search in memory involves tracing out the activation in parallel along the links from the node of each concept specified by the input words. The spread of activation constantly expands, first to all the nodes linked to each first activated node, then to all the nodes linked to each of these nodes, and so on. At each node reached in this process, an activation tag is left that specifies the starting node and the immediate predecessor.[112]

In spite of its initial appeal in many areas of research, from neuropsychological studies of semantic dementia (e.g. Warrington, 1975) to concept differentiation during development (e.g. Keil, 1979), Quillian's model is confronted with multiple issues in explaining results from psychological findings as well as with substantial theoretical issues, such as the reliance of storing knowledge at the superordinate category level rather than with individual concepts. One fundamental problem of models based on local representations plus spreading activation is that the explanation of how meaning is learnt and of how new sentences are understood is not straightforward. It seems reasonable to suppose that people understand a new sentence by retrieving stored information about the meaning of words, and then combine and alter these retrieved meanings to build up the meaning of the new sentence. In order to do that, we either turn to a set of combination rules (as partially proposed by Quillian himself) or we need to come up

[112] Much more could be said about the mechanism of spreading activation and its application in many areas of linguistic research, especially as a mechanism of word retrieval for speech production. An interesting extension of Quillian's theory was proposed by Collins and Loftus (1975), who added several processing and structural assumptions in order to deal with experimental results on semantic memory. As for speech production, for example, Dell (1986) combined spreading activation retrieval mechanisms with assumptions regarding linguistic units and rules to account for facts about speech errors; Roelofs (1992) presented a spreading activation theory of conceptually driven lemma retrieval - the first stage of lexical access in speaking, where lexical items specified with respect to meaning and syntactic properties are activated and selected.

with a different account of how the network can understand new meaning and learn it. This will be one of the main issues taken on by distributed accounts of linguistic meaning.

Allow us a final remark before turning to the discussion of neural nets: the next two sections, which aim at presenting the implication of connectionism and parallel distributed representations for understanding the nature and structure of the mental lexicon, have not been written by experts in the field and, therefore, present all the limitations and approximations typical of a sketched presentation. At this stage of our research about the structure of language within and outside the mind our aim is not to endorse a specific view or to make any contribution to existing models and we do not claim to have a complete or definitive understanding of the forthcoming issues. Nonetheless, the contribution of connectionism to the study of language and of cognition in general is definitely too important to simply gloss over it.

4.3.3. Distributed representations and connectionist models of language

The way in which we access our memory is highly multifaced and, at this point of the discussion, we can safely say that understanding and modelling memory as a storage of literal copies of items that can then be retrieved, as in a filing cabinet or in a typical computer database, does not take us any closer to figuring out the nature and structure of human memory, and, hence, of human memory for words. For example, people can recall items from partial description of their content, and they can do this even if some parts of the partial descriptions are wrong. These and other characteristics of human memory are very hard to implement in a computer, because computers store each item at a particular address, and to retrieve an item they must know that address. But the nature of human memory is surprising only because it conflicts with the standard computer metaphor.

A different landscape arises with connectionist, or parallel distributed, models. Historically, connectionism can be traced back to a tradition that is different from the one that for many decades endorsed a symbolic view of cognition and it now constitutes a valid alternative to

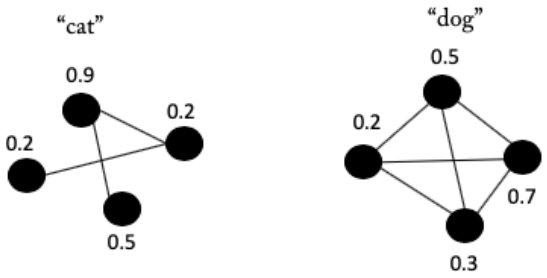


Figure 14. Example of distributed representation. For both "cat" and "dog" the units are the same, but they vary in the values and pattern of activation and in the connections between them. In order to know which concept is represented, we have to look at the entire set of nodes (Adapted from Jeffrey et al., 1996, Figure 2.16, p.91).

the symbolic model of thought.[113] As we have previously mentioned, connectionist models have a "neural inspiration" (cf. Christiansen & Chater, 1999), i.e. they are based on the idea of reproducing the fact that the brain consists of a very large number of simple processors, neurons, which are densely interconnected into a complex network, and which operate simultaneously and co-operatively. Of course, connectionist nets are not realistic models of the brain, neither at the level of processing units, which are highly simplified, nor in terms of network structure, which typically bears no relation to brain architecture.[114]

[113] Mainstream computer science arises from the intellectual tradition of viewing human thought as a matter of symbolic processing. This tradition can be traced back to Boole (1854), it runs through Turing (1937) and it finds its most substantial support in von Neumann's motivation for the design of the modern digital computers. On the contrary, connectionism can be traced back to a tradition that attempted to design computers inspired by the structure of the brain (McCulloch & Pitts, 1943). Gradually the symbolic paradigm gained ground and connectionism was largely abandoned, particularly in view of the limited power of the then current models. Now that some of the limitations have been overcome, the possibility for connectionism to be a valid source of insights on the mechanisms of cognition has been largely re-opened (cf. Christiansen & Chater, 2001b; Elman, Bates, Johnson, Parisi, & Plunkett, 1996).

[114] The fact that neural nets are inspired by the neural architecture does not mean that they are necessarily a theory of how language in organized in the brain. Nonetheless, neural networks hypotheses have been proven useful in explaining some characteristic features of the language in brain and in accounting for many neuropsychological data. The biological basis of language, specifically the anatomical specialization of different cortical areas for different language functions, is not easy to tackle onto with experimental protocols. Although some progress has been made using neuropsychological and neuroimaging methods to identify the specific relationships between cortical areas and language functions, one main finding has been that these relationships can be rather variable across individuals, suggesting that the kind of general-purpose learning mechanism hypothesized by connectionist models plays an important role in shaping the biological substrates of language. Furthermore, in neural networks specialization is often a matter of degree. Similarly, although some distinctions between functionalization of brain areas, for example between Broca's and Wernicke's areas, they fall short of explaining all of the neuropsychological data. For example, the neural basis of semantic representations is a very complicated and contentious issue, but one that neural networks

In parallel distributed approaches we can view memory as a process that constructs a pattern of activity which represents the most plausible item that is consistent with the given cues; a set of simple processing units are capable of supporting a large number of different patterns (Anderson, 1977; Hinton, 1981; Hinton et al., 1986). This is achieved by constructing representations of information in the network that are not local but distributed, i.e. they are distributed patterns of activation. In distributed representations, "[e]ach entity is represented by a pattern of activity distributed over many computing elements, and each computing element is involved in representing many different entities" (Hinton et al., 1986, p. 77) (see Figure 14). Let us just first briefly mention three important features of distributed representations that allow to model fundamental aspects of human cognitive behavior: their character is essentially constructive, they are able to generalize automatically, and they tune to changing environments (Hinton et al., 1986, p. 79). Distributed representations automatically give rise to generalizations, a helpful phenomenon in dealing with situations that are similar but not identical to previously experienced situations.

> People are good at generalizing newly acquired knowledge. If you learn a new fact about an object, your expectations about other similar objects tend to change. If, for example, you learn that chimpanzees like onions you will probably raise your estimate of the probability that gorillas like onions. In a network that uses distributed representations, this kind of generalization is automatic. […] If you subsequently learn that gibbons and orangutans do not like onions your estimate of the probability that gorillas like onions will fall, though it may still remain higher than it was initially. Obviously, the combination of facts suggests that liking onions is a peculiar quirk of chimpanzees. A system that uses distributed representations will automatically arrive at this conclusion […]. When the system learns a new fact about chimpanzees, it usually has no way of knowing whether the fact is true of all apes or is just a property of chimpanzees. The obvious strategy is therefore to modify the strengths of the connections emanating from all the active units, so that the new knowledge will be partly a property of apes in general and partly a property of whatever features distinguish chimps from other apes (Hinton et al., 1986, pp. 82–84).

The fundamental aspect of neural networks based on parallel distributed processing is that they learn; distributed representations give rise to powerful and unexpected emergent properties. Compared to local representation, distributional representations have also proven to be much more

---

have made important contributions to (for an in-depth discussion of the biology of language and connections models cf. O'Reilly & Munakata, 2000).

efficient at creating new representations: the system does not need to make a one-or-none decision on when to create a new representation, since the new pattern for the new representation can be developed over time as a result of little modifications on many separate occasions.

What does this have to do with language, semantic knowledge, and the mental lexicon? For example, Hinton (1981) demonstrated that individual prepositions could be stored in the network by adjusting the weights to make the patterns representing the proposition stable. Furthermore, several propositions could be stored in the network, in the same set of weights, and new proposition could be added without requiring additional computational elements.

Neural networks models have been able to account for a variety of phenomena (for a useful review cf. Christiansen & Chater, 1999, 2001a) in phonology (e.g. Gasser & Lee, 1990), morphology (e.g. Rumelhart & McClelland, 1986), spoken word recognition (e.g. McClelland & Elman, 1986), written word recognition (e.g. Rumelhart, McClelland, & Group, 1986; Seidenberg & McClelland, 1989), and speech production (e.g. Dell, 1986; Stemberger, 1985). If we generalize these findings at a theoretical level, models based on parallel processing of distributed representations, supported by self-organizing learning mechanisms[115], seem plausible models of learned cognitive abilities, including many aspects of language processing.

The value of connectionist approaches to language lies in the facts that: in the first place, connectionist nets typically learn from experience (see note 115 for a brief description of how feedforward networks learn through backpropagation), rather than being fully prespecified by a designer, whereas symbolic models like the ones discussed before are typically fully prespecified and do not (implicitly or explicitly) learn. In the second place, as we have seen, connectionist

[115] It seems useful to clarify some technical aspects of connectionist neural nets. The most popular of connectionist networks is the *feed-forward network*. In this type of network, the units are divided into "layers" and activation flows in one direction through the network, starting at the layer of "input units", and finishing at the layer of "output units". The internal layers of the network are known as "hidden units". Feed-forward networks learn from exposure to examples, and learning is typically achieved using the back-propagation learning algorithm (Rumelhart, Hinton, & Williams, 1986). The output obtained is compared against the correct "target" value and the difference between the two is calculated for each output unit. The squared differences are summed over all the output units, to give an overall measure of the "error" that the network has made. The goal of learning is to reduce the overall level of error, averaged across input/target pairs. Back-propagation is a procedure which specifies how the weights of the network (i.e., the strengths of the connections between the units) should be adjusted in order to decrease the error. Training with back-propagation is guaranteed (within certain limits) to reduce the error made by the network, but it does not guarantee to reduce the error to 0. One of the interesting aspects of the back-propagation principle is that it is a general-purpose learning mechanism, which has the potential to account for many aspects of language learning and, hence, to extract semantic structure and knowledge from experience, reducing the need for nativist accounts of semantic content (for an in-depth discussion of modularity and nativism in a connectionist perspective cf. Elman et al., 1996; Rogers & McClelland, 2004).

models are able to generalize. In the third place, connectionist models are able to deal with language "quasi-regularities", regularities which usually hold but which admit exceptions. In the symbolic framework, quasi-regularities were captured by symbolic rules, often associated with explicit lists of exceptions. As we have mentioned for the case of morphology, symbolic processing models incorporate the distinction between regular and quasi-regular morphology by having separate mechanisms for regular and for exceptional cases (e.g. Jackendoff, 2002). In contrast, connectionist nets may provide single mechanisms which can learn general rules and their exceptions.

The core aspect of distributed representations and parallel processing, when it comes to understanding the nature of the mental storage, is that we do not need to, and we cannot, decide upfront which information needs to be encoded and which needs to be left out; in other words, we do not need to take a stand on whether only unpredictable facts about language need be stored and predictable elements can be structured online, or if both predictable and unpredictable elements need to be stored in some kind on consultable list. As we have seen, formal serial approaches to the lexicon were driven by two fundamental assumption. First, the lexicon was assumed to be a compositional derivational system. Complex words were believed to be generated from simpler forms. Second, the set of atomic elements was assumed to comprise any word or formative that is not predictable by rules. Ruled-governed combinations of these atomic units, the regular complex words, were assumed not to be available as units in the lexicon, as their storage would introduce unnecessary redundancy in the model. Instead, connectionist networks have the advantage that they can account for graded, probabilistic phenomena (for an in-depth discussion of storage and computation in the mental lexicon cf. Baayen, 2007). In connectionist models, representations are not stored in a discrete manner, but are rather built by a process of reconstruction and integration based on the information distributed in the system that allows to focus on the actual richness of linguistic structure and language processing. Neural network models allow for more parsimonious accounts of the often complex web of regularities and exceptions in language by modeling them with a unified set of principles (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg, 1997). In comparison to localist network approaches, connectionist approaches seem to get closer to grasp the open-ended nature of language and the high dimensionality required for language processing (Elman, 1991). Models which have used the localist approach have typically made an *a priori* commitment to the nature of linguistic representations and networks are then explicitly

trained to identify these representations. In the alternative approach, distributed representations and learning algorithms do not depend on preexisting preconceptions about what the abstract representations are.

One of the downsides of distributed representations is that they are very difficult to analyze, i.e. to understand what kind of information is encoded in the units and how it changes when the learning algorithm modifies the node or the weights of the edges of the network. Various techniques exist to analyze distributed representations, like cluster analysis (e.g. Elman, 1990), but the results of experiments targeting the internal structure of distributed representations is overall limited (Elman, 1991). This means that, compared to previous model descriptions, we are not able to say what kind of information is stored in the basic units; the upside is that we do not need to. Connectionism should not be viewed as an additional method for studying language processing, but as a new approach that offers an alternative to traditional theories, which describe language and language processing in symbolic terms. In fact, Rumelhart and McClelland suggest

> that implicit knowledge of language may be stored among simple processing units organized into networks. While the behavior of such networks may be describable (at least approximately) as conforming to some system of rules, we suggest that an account of the fine structure of the phenomena of language and language acquisition can best be formulated in models that make reference to the characteristics of the underlying networks (Rumelhart & McClelland, 1987, p. 196).

The underlying theme is that by studying the networks, their structures, and their mechanisms, we may get some useful insights on how cognition in general, and language processing in the specific case, works and is organized. According to this view, connectionist models provide the starting point for a theory of semantic cognition (as, for example, the one proposed by Rogers & McClelland, 2004); in the first place, performance in semantic tasks occurs through the propagation of activation among simple processing units, via weighted connections. In the second place, connection weights encode the knowledge that determines which distributed representations arise internally, and that governs the joint use of these representations to determine the outcome of the process. In the third place, the connection weights are learned, so that semantic cognitive abilities arise from experience.

A review of the variety of connectionist models that have more or less direct implications for a distributed account of the mental lexicon goes far beyond our possibilities, needs, and scopes.

The aim of the present section was to propose an informal account of those principles underlying connectionist approaches that have shown to be relevant, suitable, and efficient in modeling and explaining linguistic behavior. At the current state of affairs, much is still left to do both in computational neuroscience and in psycholinguistic approaches to justify and further develop distributed approaches to cognition in general and, in our specific case, to linguistic meaning at a biological as well as psychological level. Connectionism has already had a considerable influence on the psychology of language. But the final extent of this influence will depend on the degree to which practical connectionist models can be developed and extended to deal with complex aspects of language processing in a psychologically realistic way. The future of connectionist psycholinguistics is therefore likely to have important implications for the theory of language processing and language structure, either in overturning, or reaffirming, traditional psychological and linguistic assumptions. In the next section, we will review an interesting development of connectionist account of meaning, i.e. the possibility that the very idea of the mental lexicon may be wrong and that we do not actually need any kind of word storage in the mind in order to process language. To get to the core point, we will need to describe one specific network model, Elman's *Simple Recurrent Network* (SRN), the discussion of which will allow us to have a better practical grasp on how and what networks learn.

3.3.4. Do we actually need a mental lexicon?

One of the main issues of early neural nets was how to represent time, i.e. to provide networks with a dynamic memory. It goes without saying that accounting for time is fundamental to model language processing, given that language happens in a temporal sequence. In order to understand how this has been possible we need to introduce an aspect of connectionist networks we have glossed over in the previous section: hidden units. One of the basic principles which drives learning in neural networks is *similarity*[116]: similar inputs tend to yield similar outputs. This principle allows networks to generalize their behaviors beyond the cases they have already

[116] In a network framework, similarity has to be understood in terms of similarity of *input patterns*, which can, in turn, be thought of as points in a representational space, i.e. as vectors. The similarity between vectors is their Euclidian distance in the representational space. In technical terms, the problem we are discussing arises when patterns which are distant in space must be treated as though they were similar. We will come back to the technical definition of *similarity* in the chapter concerned with computational implementations of distributional semantics (cf. chapter 5 and paragraph 5.3.2).

encountered. Nonetheless, the principle of learning by similarity opens as many questions as it resolves; in fact, in many cases we are faced with inputs that may resemble each other superficially, but that instead have to be treated in different ways. Examples of this sort in language are countless: a child learning English needs to figure out that, for example, *bake* and *make* form a different past tense even if they look very similar. Even more problematic for a language learner is the situation in which the exact same word means very different things and can be used in very different ways (cf. Elman, Bates, Johnson, Parisi, & Plunkett, 1996). The issue underlying these examples is that in order to learn we need more information that the one provided by similarity.

A way to overcome this issue is to add a layer of nodes between the input layer and the output layer; the units of this extra layer are generally called "hidden units" and can be thought of as equivalent to the internal representations we invoke in psychological theorizing. Allowing the network to build internal representations in the hidden layer makes networks very powerful: networks learn how to treat similar inputs as different by transforming the input representations to a more abstract kind of representation.

This is the basic architecture of feed-forward networks: the flow of information proceeds from the input through successive layers and culminates in output units (cf. footnote 115). The flow of information is unidirectional, from input to output, and at the conclusion of a process of analysis of the input, all previous activations are wiped clean and the next input is received and processed. Not exactly how language processing might work. In these early networks, processing was essentially atemporal, in that the changes of the weights implemented a semantic long-term memory, reflecting the time course of learning, but they lacked a short-term or working memory (cf. Elman et al., 1996, p. 74). An implementation of short-term memory is possible by allowing connections from nodes back to themselves; in this way, the network's activity at any point in time can reflect whatever external input is presented to it and the network's prior internal states. Such networks are called *recurrent* networks. Recurrent networks have fully interconnected hidden units; each hidden unit activates itself and the other hidden units. As a result, prior acquired knowledge stored in the hidden units is activated every time an input is processed (for a very popular recurrent model cf. Elman, 1990). In this kind of recurrent networks, time is represented

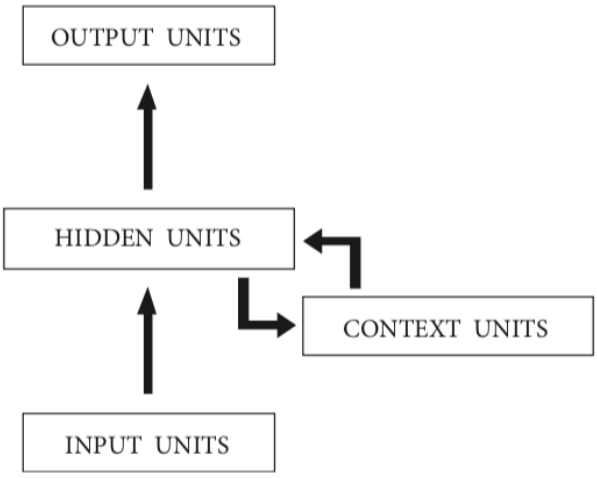


Figure 15. Elman's Simple Recurrent Network architecture. Information flows from input to hidden to output layers. In addition, at every time step t, the hidden-unit layer receives input from the context layer, which stores the hidden-unit activations from time t–1.

– short-term memory is given to the network - implicitly by allowing it "to be represented by the effect it has on processing. This means giving the processing system dynamic properties that are responsive to temporal sequences" (Elman, 1990, p. 182).

Elman's SRN (Elman, 1990, 2004) is a network designed to process events and behaviors over time, in which the key to the architecture is recurrence. The network is composed of an input layer, an output layer, a hidden layer, and context units (see Figure 15 for a schematic representation). Outputs are driven by the internal ("mental") states, patterns across the hidden units, which are themselves the product of the input layer and of the context units, which are the network's own previous states. The network starts with no previous knowledge and the weights of the system are only initialized with small random values, i.e. no information is hard-coded in the net. The SRN is fed with a large number of inputs and outputs. Given its architecture and a large amount of training data, the network learns not only to replicate a certain output given an input, but to generalized beyond already encountered sentences and, furthermore, to predict the next word (or set of words) that might follow a given input. Rather than memorizing the corpus, the network learns to predict, in a context-appropriate manner, all the words that are grammatically possible, with activations levels in the network corresponding to the probability that each word might occur. By looking at the "mental" states of the network in the hidden units, Elman was able to show that the network learned how to induce categories such as Noun, Verb, or Animate (that the patterns evoked in response to each word did reflect the word's category membership), as well as

categorizations at lower levels of granularity (e.g. transitive, intransitive, optionally transitive verbs) – starting from training data only.

Another aspect of Elman's SRN that makes it interesting from the perspective of studying linguistic meaning is the fact that in the network as well as in language there are no single abstract representations that might reflects what is known about a word. Each occurrence of a word results in a state that is similar to all other occurrences, but not identical (i.e. it inhabits the same space region of the other members of the lexeme, but they all have slightly different representational vectors). One of the fundamental aspects of this family of network - for what concerns a definition of the mental lexicon - is that it can retain information that is highly word-specific as well as context-dependent and that it can do it without needing to hypothesize stable and fixed representations. The network learns to make context dependent predictions thanks to the fact that its internal representations reflect both grammatical categories and lexico-semantic information in a dynamic way. Rule-like behavior is achieved through the system's dynamics; the trajectories or paths of activation over time though the network's internal state continuously change and adjust in accordance to the unfolding of the input. These trajectories reflect the intrinsic dynamics of the network, such that only some paths through state space are felicitous and well formed (Elman, 2009, 2011). The network dynamics encode what we would conventionally think of as the grammar that underlies the language sample and it is directly dependent on the words in the language sample.

> The critical insight here is that the role of words in such a dynamical system is to function as external stimuli that alter the system's internal state. The effect that a given word produces is a function of two things: the prior state of the network, which encodes the context in which word input occurs; and the network's dynamical structure or grammar, which is encoded in its weights. In this scheme of things there is no data structure that corresponds to a lexicon. There are no lexical entries (Elman, 2009, pp. 555–556).

The insights provided by SNR and similar networks, together with data on the networks' performances in predicting how humans deal with complement structure preferences, or with verb specific thematic roles, or many other aspects of everyday language processing, deeply challenge the idea and possibility of a mental lexicon, and especially of the mental lexicon as a dictionary. In fact, if SRN had a mental lexicon, then it would have stored a lot more information that what previously envisaged. Lexical knowledge is indeed dynamic, context dependent, and event

knowledge is shown to play a significant role in early stages of sentence processing and structural analysis. In addition to this, the long-standing issue of sentence processing as happening in real time, incrementally word-by-word, can be observed from a different perspective when analyzed in the framework of the systems described above.

As we have seen, within the psycholinguistic literature, much of the data that led to new views about the lexicon have resulted not from the direct study of lexical representations, but as a by-product of theoretical standpoints regarding the mechanisms of sentence processing. One of the main problems has always been the fact that sentences are at least temporarily ambiguous in the sense that they are compatible with very different grammatical structures and very different meaning interpretations. Traditional two-stage processes do not seem to provide convincing frameworks or computational explanations, at least at the current state of the research. One of the reasons why two-stage processing theory seem to fail is the definition of mental lexicon they endorse; taking the mental lexicon as a limited dictionary storage of phonological, morphological, syntactic, and semantic content would hardly allow to account for the potentially unbounded number of contexts that might be relevant for understanding a word or a sentence.

A different view of the mental lexicon is proposed by Elman (2004, 2009, 2011), in which words do not *have* meaning, but they rather provide *cues to meaning*.

> But one can imagine thinking about words, what they mean, and how they are stored, in a very different way. Rather than putting word knowledge into a passive storage (which then entails mechanisms by which that knowledge can be 'accessed', 'retrieved', 'integrated', etc.), words might be thought of in the same way that one thinks of other kinds of sensory stimuli: they act directly on mental states. This by no means is to deny that the nature of this interaction is complex or systematic. Indeed, it is in the precise nature of their causal effects that the specific properties of words – phonological, syntactic, semantic, pragmatic, and so forth – are revealed (Elman, 2004, p. 301).

In the framework described by Elman, saying that words "act directly on mental states" or that they are "cues" to meaning can be interpreted in the sense that words trigger representations in the "mental" space that are not words, or part of words themselves, but rather patterns of activation in the hidden layer that depend of previously processed stimuli and are, therefore, context and event dependent. The core ideas that motivate this proposal have deep roots in the cognitive as well as linguistic literature; for instance, the meaning of a word is grounded in our knowledge of both the material and the social world; words never come without a context; prediction is a well-known and

powerful engine for learning, as well as an highly adaptive mechanism; events play a major role in organizing our experience (for a discussion cf. Elman, 2011 and bibliographic references therein).

The word-as-cues approach suggests that: in the first place, we do not need to decide what ends up being stored in the mental lexicon because language processing is not a matter of retrieval but of activation and continuous updates of our internal states. In the second place, the integration of various linguistic as well as non-linguistic aspects is the starting point for any possibility of theorizing and modeling human language abilities. In the third place, such a view of the mental lexicon suggests that linguistic knowledge is indeed acquired and constantly depends on our experience of the world, i.e. not only we learn words from our linguistic experience, but we also learn syntactic patterns and linguistic structures from positive data only. Needless to say, a lot of work still needs to be done to go from these general principles and assumptions to psychological and computational models of how language works and is structured in cognition. Nonetheless, these renovated accounts of the mental lexicon and of language processing opened the way to fervent debates and innovative paradigms, some of which we will come back to when discussing the computational implementations of the distributional semantic hypothesis.

We have come a long way from when we opened the chapter saying that we would not have been able to say much about the nature and structure of the mental lexicon. There is clearly a lot more to say about this topic that at a first superficial glance. As we have pointed out multiple times throughout the chapter, one of the main reasons for which the study of the mental lexicon is such a wobbly territory is that the very definition of the object of inquiry is often only the by-product of specific paradigms of research, interested in the nature and structure of human processes more that on linguistic meaning itself. We have merely flown over the intricate maze of approaches, theories, and models that over the past century contributed to the theoretical and empirical journey of how language is structured in the mind. And our journey is not over, yet. One other way in which the mental lexicon has been largely approached and studied has been by means of free association tasks. The reason that led us to consider free word association in a separate section is that methods and observations that are based on free association data seem to be primarily concerned with the organization of words in the mind, i.e. of how words are linked to each other in the mental lexicon, and only in a second moment interested in evaluating the mechanisms of production and comprehension.

4.4. Free word associations and semantic networks

So far, we have been mostly dealing with accounts of language processing and of the mental lexicon which assumed that, in order to process language, we need to view linguistic information as, in one way or another, decomposed into either its various constitutive parts (phonetic, morphological, syntactic, semantic), or as a distributed pattern over a connectionist network. In the first case, conclusions about the structure of the lexicon were mostly drawn from strong theoretical assumptions about the independency of linguistic information from the rest of our knowledge about the world, the modularist view of our cognitive architecture, and the partitioning between rules and content, on the line of generative views of language. In the second case, distributed views of word meaning were grounded on biological observation about the nature of representations in the brain and on the assumptions and characteristics of parallel distributed networks. Nonetheless, if the take the perspective on any average language user we would intuitively say that we usually deal with words, compounds, sentences, texts and so on. Such intuitions are backed up by research paradigms based on data gathered by means of the well-known psycholinguistic method of *free word associations*. In the next few pages, we will briefly outline some of the main reflections and outcomes based on associative data, what they are, and what use researches have been making of them.

A preliminary note is here in order. Free word associations became a relevant method within approaches stemming from the second-cognitive revolution. The importance of how words are connected and elicited in our minds is an outcome of the shifts in perspectives in studying meaning within cognitive semantics. As we have seen before (§ 3.5.1.), second-generation cognitive linguists take a pragmatically enriched view of meaning modeling, where natural language use is of key importance (cf. Croft & Cruse, 2004; Fillmore, 1975; Geeraerts, 2010; A. E. Goldberg, 2006; Paradis, 2005). Lexical items are not taken to have stable meanings, but rather to evoke meaning in context-dependent ways. Meaning creation in context in both dynamic and constrained by encyclopedic factors and conventionalization patterns. Cognitive semanticists highlight that the way in which people understand language is related to the world of experiences that surrounds them. This – together with other fundamental insights in the study of meaning - are variously addressed by looking at how people connect aspects of their experiences by means of looking at how the lexicon is organized. Given the connection established by cognitive

semanticists between concepts and words, free word associations are often taken as a window on our linguistic/conceptual structure.

4.4.1. The method, early results, and shortcomings

In strictly methodological terms, the free association task is extremely simple: a person is asked to write down the first word(s) that spontaneously come to his or her mind after reading a cue word, with no restriction about the type of answer that the person can produce. Less straightforward is the interpretation of the results. If associations between words actually offer a window on structures and processes in the human mind, then how do we go from having a bunch of association data to a whole theory of the structure of the mental lexicon?

The first recorded experiment on the organization of words in the mind by means of word associations is Francis Galton's experiment on himself: he wrote down 75 words of slips of paper and after a few days recorded his response times in writing the first two ideas that came to his mind after glancing at the words he had previously written and forgotten about (Galton, 1880). The replications of Galton's study on more and more subjects showed that the associations for the cued words people tend to come up with are not totally random: people tend to give rather similar responses, so much so that one can talk about "norms of word association". For example, people tend to associate *low* to *high, black* to *white*, *king* to *queen*, with such a high frequency that researches started to think that the frequency of a response was likely to be a good index of its strength (for one of the first collections of norms of word association cf. contributions to Postman & Keppel, 1970)[117].

[117] Norms of free words associations available nowadays are: the largest British English word association databases are the Edinburgh Associative Thesaurus (Kiss, Armstrong, Milroy, & Piper, 1973) and the Birkbeck norms (Moss & Older, 1996). For American English, the most commonly used dataset is the University of South Florida dataset (Nelson, McEvoy, & Schreiber, 2004) and the Small World of Words dataset (De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2018). Datasets in other languages are available for Dutch (Dutch Small World of Words: De Deyne, Navarro, & Storms, 2013; De Deyne & Storms, 2008b), Korean (Jung, Na, & Akama, 2010) and Japanese (Joyce, 2005). The Small World of Words (SWOW) dataset for both American English and Dutch is the most extensive set of word associations available to date. A particular feature of the SWOW is that it is based on a multiple-response rather than on a single-response procedure (participants can list up to three associations and the strength of each association is calculated on the base of the order in which associations were listed). the multiple-response procedure results in a more heterogeneous set of responses, which lead to better predictions of lexical access and semantic relatedness than do single- response procedures (De Deyne et al., 2013).

Norms of word associations may give us an idea of how words are linked together in a person's mind. Firstly, the notion of associative strength is important in the explanation of several psychological phenomena such as our ability to recall words from episodic memory (Nelson, Schreiber, & McEvoy, 1992) and semantic priming (Lucas, 2000) - a task in which the recognition of a target word is enhanced by an earlier presented cue word that is strongly related in contrast to a weakly related one. The consistency of the associative responses, hence the strengths of the connection between pairs or set of words, together with their applicability in explaining psychological data, gave rise to the idea that it would have been possible to draw a map of the mental lexicon based precisely on responses from association tasks.

Unfortunately, one serious shortcoming of free words associations is that they indeed tell us with which words a given word is connected, but they do not tell us what kind of connections are entailed between words, i.e. the actual structure and types of the links between the words.

Aitchison (1987) offers a clear review of some types of links that can be individuated starting from associative data and building on top of the traditional distinctions between semantic relationships. The commonest link between cue and response words seems to be *co-ordination*: links between co-ordinates, like *salt* and *pepper*, *left* and *right*, *hot, cold, warm* and *cool*, i.e. between words that cluster together and the same level of detail, seem to be the strongest kind of links (i.e. the one that occur with the highest frequency). Data on the strength of co-ordinate relations are supported by research on both semantic slips of the tongue as well as on aphasic patients. In the first place, errors from slips of the tongue and word blending show that words are stored in semantic fields and that co-ordinates are closely associated. In the second place, aphasic patients often produce a co-ordinate or close relative to the target and similar types of mistakes are made in comprehension too. Aitchison draws the following conclusions on the nature and organization of co-ordinates in the mental lexicon: "[i]t is difficult to be precise about the detailed organization of co-ordinates within the mental lexicon, since the structure of a group is likely to depend on the type of word involved - objects, colours and actions might be treated rather differently. It seems probable, however, that for each group there is a nucleus of closely linked words, with other words attached somewhat more loosely round the edges" (Aitchison, 1987, p. 77).

The second strongest link is *collocation*; the next most common response, after co-ordinated words, involves a word which is found together with the stimulus in connected speech.

Collocations links cover a wide spectrum of possible cases, from words which are optionally, but commonly associated (*rude adolescent*, *unruly hair*), to habitual connections or clichés (*agonizing decision*), to idioms which overall meaning cannot be predicted by the sum of the individual words (*fall into place, call it a day*).

At the third and fourth place for association strength we find hierarchical structures and synonyms. As for hierarchical structures we have already discussed Collin and Quillian's (1969) proposal and experiments in the previous section. To summarize the finding about the hierarchical organization of the mental lexicon, superordinate levels may be easily available when the contents of a group are fairly prototypical, and the superordinate level is a commonly used one. But the vertical movement in the hierarchy seems to require more effort than horizontal movement among co-ordinates. With Aitchison, "[l]inks between a word and its superordinate were firm when the superordinate was a frequently used term and when the words came from the same general meaning area, such as bird which included robin, thrush, and so on. All these findings support the suggestion that words from the same semantic field are closely linked. In contrast, bonds between words from different semantic fields were somewhat weaker" (Aitchison, 1987, pp. 83–84).

Hence, the unconstrained nature of the free association task makes difficult to grasp what associative responses tell us about the lexicon. Another way to go about identifying the nature of the responses is to investigate whether the cue-response relationship is paradigmatic or syntagmatic. In a syntagmatic relationship (*smelly-cheese*), cue and response have a different syntactic role in a sentence and can belong to different part-of-speech classes. Syntagmatic relationships are the equivalent of the relationship of collocation in the above classification. In a paradigmatic relation (*cheese-socks*), words have the same syntactic function in a sentence.[118] De Deyne and his collaborators (2015) suggest that in the mental lexicon as represented by free associations there seems to be a dominance of paradigmatic responses for nouns (especially for concrete nouns) and syntagmatic responses (mostly nouns) for adjectives and verbs. A second possible, and very often used, distinction is the one between taxonomic and thematic[119] responses;

[118] It might be useful to point out how extremely simplified is the definition of paradigmatic and syntagmatic relationships when compared with the one originally formulated by De Saussure. The distinction between paradigmatic and syntagmatic seems to be reduced to whether two associated words belong to the same part-of-speech class. Needless to say, in the Saussurian framework, the two ideas were definitely more complex and articulated – e.g. not all nouns form a *paradigma* and not all adjective-noun pairs are (or can be) in a syntagmatic relationship.

[119] The distinction between taxonomic and thematic relationships is of great interest in understanding the structure of the lexicon. Nonetheless, there is very little agreement on what distinguished a taxonomic from a thematic relationship,

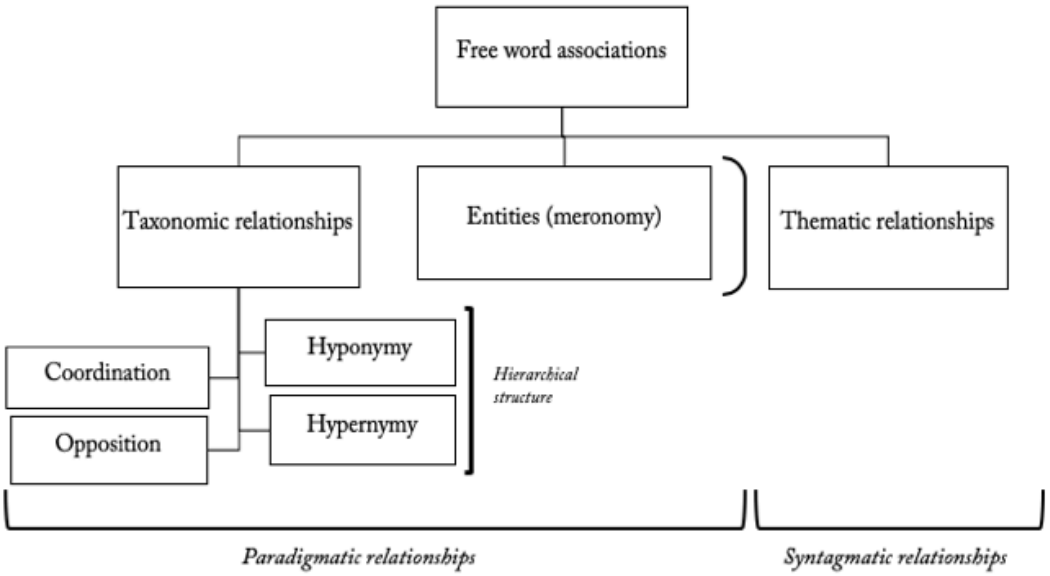


Figure 16. Schematic of a possible organization of the types of associations according to the reviewed recent literature. It is interesting to note how all these factors are present in free word associations, suggesting that the meaning of words is in fact quite rich and multifaceted.

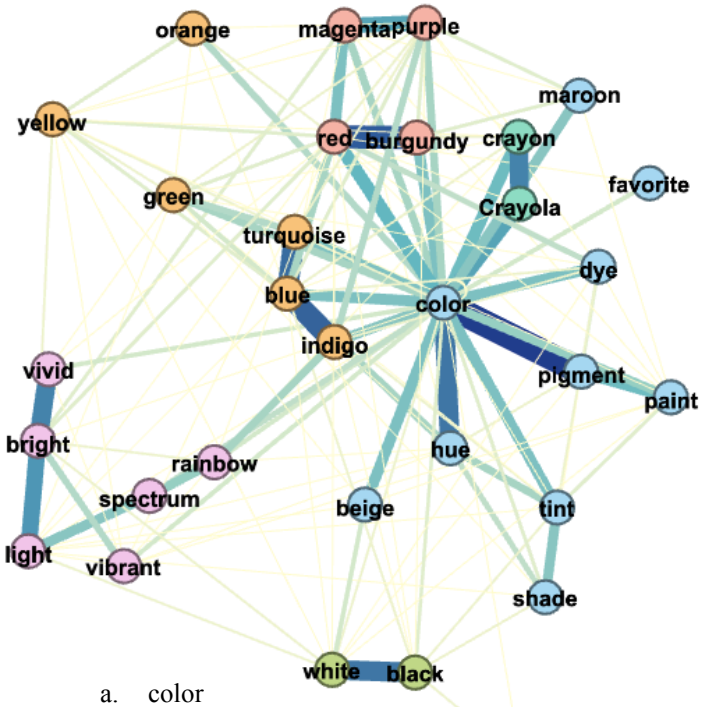
taxonomic responses include contrast, coordination, superordinate and hyperonymy, subordinate or hyponymy responses. Meronomy responses (called *entity*) correspond to attributes or features that are part of the concept (such as *dog* and *tail*). Drawing on the distinction between word's denotation and connotation, the authors suggest that

at least in the psycholinguistic literature. In order to properly address the issue and to review the existing literature we would need to write another chapter completely dedicated to the topic. Since this would take us even further away from our original goal, i.e. discussing the organization of words in the mind, we will limit ourselves to some preliminary considerations, with the hope to provide a more in-depth discussion in further studies. If we take the notion of taxonomic relationships in its broad sense, taxonomic relationships are semantic relationships of the paradigmatic axis. The hierarchical structure of our knowledge has been one of the main focus of research; for example, Rosch (1973) argued for a hierarchical taxonomic structure of natural categories and we can recall Collins and Quillian's (1969) study on the hierarchical structure of the mental lexicon. Thematic relationships are more difficult to define since the notion of thematic cannot be taken as overlapping with the notion of syntagmatic relationship. A useful working definition of thematic relations for the purposes of our discussion is the one provided by Lin and Murphy, according to which "thematic relations are the external or complementary relations among objects, events, people, and other entities that co-occur or interact together in space and time" (Lin & Murphy, 2001, p. 3). Some examples of thematic relationships are spatial (e.g. *roof-house*), functional (e.g. *chalk-blackboard*), causal (e.g. *electricity-light*), and temporal (e.g. *restaurant-bill*).

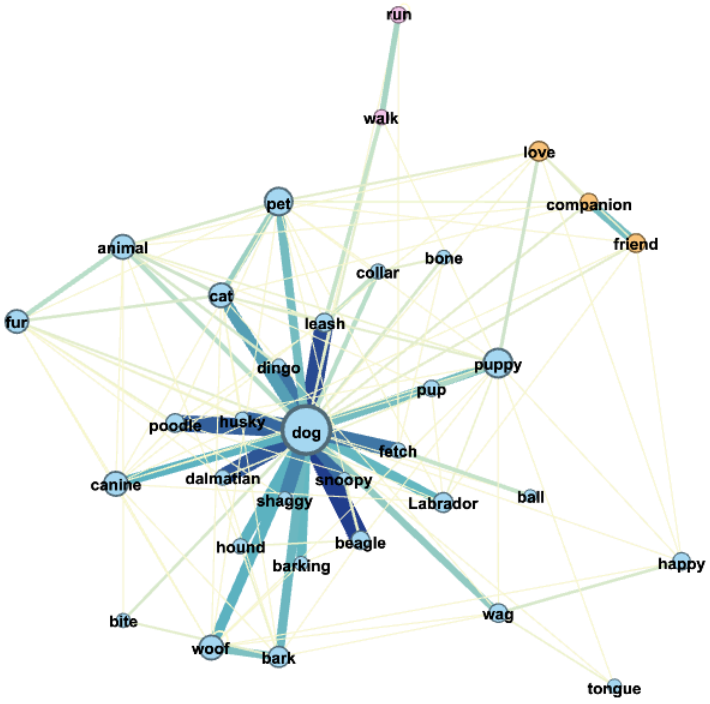
> [w]hile both entity and taxonomic responses contribute to what is often understood as the denotation of a word or the semantic content necessary for defining a concept, the most common responses in word association data are those that capture thematic relations, indicating some sort of contiguity in time and space rather than similarity in meaning (De Deyne & Storms, 2015, p. 6).

Apart from the pitfalls of such formulation, we can see how it clearly points towards the direction of an encyclopedic view of meaning. Free association data could hardly support a view in which linguistic meaning is strictly separated from non-linguistic meaning, as the dictionary view of meaning would suggest. Free word associations seem to tell us that the way in which we construct and organize our mental lexicon, i.e. our knowledge of language, depends on our overall knowledge about the world. For example, if we had to classify some associations as taxonomic or thematic and to assign the status of linguistic relationships to the taxonomic ones and of general-knowledge relationships to the thematic ones, associations like the ones between *dog* and *bone*, or *dog* and *leash*, would seem to depend on a *frame*-like structure of our knowledge more than on a strictly linguistic structure, as would instead be the case for the co-ordinates *dog* and *cat*, or the superordinate relationship between *dog* and *animal* (for the continuation of this discussion see Box 1). Hence, both types of relationships, i.e. both kinds of information, seem to play a fundamental role in structuring our knowledge about language, suggesting that the distinction between what is linguistic and what is not may lead to profound confusions in the interpretation of the data.

Box 1. A closer look at free word association data



a. color



b. dog

The two networks represent the strengths and relationships of word forward associations for *color* (a) and *dog* (b) (forward associations are those associations that have a certain target word as cue word; backwards associations are those associations for which the target word is the responses of a cue word) taken from the En-SWOW data. In our case, we are representing the first 30 forward associations of our target word as well as data from the first 10 forward associations for each of the 30 associates of the target word. The color of the nodes indicates the communities in the network. The size and color of the edges indicate the associative strength between the nodes. Starting from what we have said so far about the possibility of establishing the type of connection between the words, there are at least some preliminary qualitative observations we can suggest by simply looking at the networks. In the first place, all the types of associations we have previously mentioned can be found in the actual data. For example, a variety of taxonomic relationships can be observed among the strongest connections: *color-blind* or *favorite-color* are typical examples of fixed phrases and word collocations (even though it has a weak connection in this representation); both *color* and *dog* elicit a wide range of subordinates; *dog* is also associated with the co-ordinate *cat* and with its hyperonyms *animal* and *pet* (we can notice that the connection with *pet* is stronger than the one with *animal*). As expected, our target words are also associated with what are generally called thematic relationships: *dog* has very strong connections with *leash* and with *fetch* (the ball). *Color* is associated with *hue, pigment, dye*, and so; it seems hard to find a semantic relationship in which these associates might comfortably fit. Even though these preliminary observations are interesting from a speculative point of view, to the extent of our knowledge, there has not yet been neither an extensive survey of the kinds of associations observable in associative data, beyond the once we have already reported, nor possible implementations of automatic detection of such relationships.

4.4.2. From associative pairs to associative networks

A further shortcoming of the initial research on free word associations was that it limited the discussion on the associative strength between isolated word pairs. According to Deese (Deese, 1965; Szalay & Deese, 1978), in order to study the meaning of words, the central issue is not the single connection but how the meaning of a word is conveyed by the entire set of connections in a larger network of knowledge. This observation shifted the focus from the stimulus-response properties, such as frequency and associative strength between two words, to their response distributions and, more broadly, to the investigation of the general, large-scale properties of the networks.

As we have seen, semantic networks, in which words (nodes) are connected with each other through a set of links (weights)[120], have been introduced by Collins and Quillian (1969) and by Collins and Loftus (1975) and have remained an influential theoretical model of the mental lexicon ever since. The idea of deriving large-scale networks from word association data has seen a renewed interest in recent years, based especially on the intuition that a large-scale representation of word associations by means of networks may offer insights on the organization of language in the mind and of how we process it not only at the microscopic level of analysis (the level of the local interactions between pairs of words), but also at the mesoscopic level (cf. De Deyne, Kenett, Anaki, Faust, & Navarro, 2017) - that resembles the level of the *local holism* in semantic field approach discussed in § 3.2. - and at the macroscopic one - which is the level of the global characteristic of the network - like network topology and centrality (cf. Borge-Holthoefer & Arenas, 2010; De Deyne et al., 2017). The main intuition that drove researchers along this line of inquiry was that networks are not only a useful way of representing large sets of interconnected data, but that, at the same time, language exhibits some fundamental properties of complex networks and, hence, its organization and behavior can be explored by means of network analysis. Especially the study of the global structure of the network has become a frontier in the field, principally thanks to the developments in network theory (Baronchelli, Ferrer-i-Cancho, Pastor-

[120] In contrast to distributed semantic networks, where the meaning of a single word involves multiple processing units, semantic networks of the kind discussed below are often constructed a localist network, where each node in the network corresponds to a word (De Deyne & Storms, 2015). In the framework of the distinction presented in the introduction to the mental lexicon as a network, associative networks of the kind discussed here are a kind of cobweb networks.

Satorras, Chater, & Christiansen, 2013; Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006; Costa et al., 2011; Newman, 2003).[121]

The analysis at the mesoscopic level is the one that interests us the most following the theoretical discussion that we have been constructing so far, given that it allows us to investigate the structure of the elements that compose the network (for example by looking at the its organizational principles and distinguishing between taxonomic and thematic relationships) and to account for a wide range of cognitive phenomena (for an in-depth review cf. De Deyne et al., 2017; De Deyne, Verheyen, & Storms, 2015), such as semantic relatedness (e.g. Borge-Holthoefer & Arenas, 2010), associative and semantic priming (eg. Cañas, 1990; Chwilla & Kolk, 2002; Hutchison, 2003; McNamara, 1992), and word retrieval processes (e.g. Capitán et al., 2012; G. W. Thompson & Kello, 2014). Focusing on the mesoscopic level of analysis allows to use associative data to obtain a useful representation of word meaning and to use networks or graphs derived from association data to learn about the development of language, its organization, and so on, i.e. to tackle onto the mechanisms of language processing.

The mesoscopic (or group level) involves the properties of a considerable subset of nodes in the network and it seems to be informative about the meaning of words. Starting from a large-scale linguistic network, it is possible to identify and compare clusters through community

[121] Inspired by the complexity theory, it has been recently suggested that language is a complex system (e.g. Beckner et al., 2009; Markošová, 2008). It has been argued that, if language is conceived of a complex system of linguistic units and their relations, it is expected to exhibit emergent properties at the system-level due to the microscopic-level interactions between the system elements. Complex network theory seems to provide appropriate modeling for human language as a complex system and powerful quantitative measures for its complexity at the system-level. For example, linguistic networks are multi-level systems which seem to exhibit small-world structure (typical of other occurring natural networks), which means that they are characterized by a high amount of local clustering but an average short path length (e.g. De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005). One possibly interesting fact in the analysis of language in term of complex networks is that it can translate into methodologies for linguistic research at every level of analysis, uncovering a series of statistical laws of human language and suggesting a wide range of possible points of connections with the methods and interests of quantitative linguistics. For example, in a linguistic network model, all the linguistic units are treated as identical in that they are all modeled as vertices. In quantitative linguistics, however, linguistic units (such as words) can be measured in terms of such indices as frequency, length, polysemy, and polytextuality; network theories are moving in the direction of incorporating these kinds of additional information in order to find out how they correlate with quantitative measures of the networks (for a review of the currest state of the research cf. Cong & Liu, 2014; Liu, 2011). Furthermore, it has been argued that language networks can be viewed as an approximation of our cognitive system, based on observations on the interplay of cognition and language as captured by a network representation and analysis (Beckage & Colunga, 2015). As far as our discussion about the mental lexicon is concerned, the current state of the research on language as a complex network seems to give us only marginal insights, given that it is currently mostly focused on the macroscopic large-scale properties of the networks. Analysis at the mesoscopic and microscopic level, which seem to be the levels which are more directly involved in the representation of the structure of the sub-systems of the lexicon, and that, therefore, allow to focus of the structural relationships between words, are briefly reviewed in the main body of the paragraph.

detection methods[122] and to investigate the size and type of these communities. Clusters are groups of nodes with a higher level of interconnection among themselves than with the rest of the network.

In the first place, the identification of clusters at the mesoscopic level allows to evaluate the principle underlying the exhibited structure. As we have seen, there are many different ways in which the mental lexicon can be structured at the mesoscopic level. One of the most influential ideas in psychological theories about knowledge representation is that of a hierarchical taxonomy, in which concepts are grouped in progressively larger categories. However, recent advances in network studies based on free associations reveal the overall absence of a basic-level taxonomy even for biological categories and the widespread thematic structure across nearly all communities, strongly suggesting that multiple factors contribute to structure in the mental lexicon, and thematic relations are a major one of them (De Deyne et al., 2015). This was made possible by the fact that the selection bias towards concrete words (and natural categories) was removed by the implementation of large-scale networks representing all types of words including adjectives and verbs (for the role of thematic representations even in domains such as animals cf. Lin & Murphy, 2001). Furthermore, observing the structure of the mesoscopic level allows to grasp qualitative properties and similarity relations between words in the lexicon. Looking at the response distribution, instead of at the single connection, opened new possibilities of research, especially in the direction of quantifying the similarity between two (or more) words. The new method introduced by Deese offered a way of quantifying how closely related two words are in meaning by looking at the commonalities of distribution between them. Two words will have similar meaning not because they are associated, but because they have many associates in common. This notion of similarity introduced by Deese inspired many modern approaches to semantic alignment that rely on distributional similarities. Statistical techniques allow us to infer that two words have similar meanings (i.e. they are semantically closely aligned) even though they are not strongly related and rarely occur in the same text. These models will be the focus of the next chapter. Nowadays, the notion of semantic similarity as the degree to which the distribution of associations or of neighbors (for distributional approaches) overlaps has seen very fruitful developments and

[122] There are various methods to identify clusters in large-scale networks. One of the most common methods is the *Order Statistics Local Optimization Method* (OSLOM; Lancichinetti, Radicchi, Ramasco, & Fortunato, 2011), which allows to identify overlapping and hierarchical communities, grouping smaller communities in larger ones and allowing to investigate different levels of abstraction on the line of hierarchical networks and taxonomy-based theories (for a recent application of the method cf. De Deyne, Verheyen, & Storms, 2015).

applications (for example, for a very recent quantification of semantic alignment cross-linguistically on the basis of distributional data cf. B. Thompson, Roberts, Roberts, & Lupyan, 2018). In addition, a further measure of semantic similarity has been proposed by Kenett and colleagues (2017) on the basis of the idea put forward by Collins and Loftus that semantic distance is the "shortest path [direct or indirect] between two nodes" (Collins & Loftus, 1975, p. 412, note 3), and implemented by means of network technologies applied to semantic networks built from free associations. As such, semantic distance can be defined as the number of steps that intervene between the prime and the target in memory. This approach seems to offer interesting predictions in how path length affects performance in semantic relatedness judgment tasks and recall from memory.

In conclusion, the microscopic or node level of analysis of the network focuses on how a single node is connected with the rest of the network. Examples of measurement at the microscopic level are node centrality measures, such as the number of in- or outgoing links. These type of centrality measures have been studied quite extensively in psycholinguistics to explain why certain words are processed more efficiently than others. In this case, network-derived measures provide a structural explanation for many lexical properties of words which have been demonstrated to facilitate word processing.

Allow us some final general remarks. Among the growing enthusiasm for complex networks in recent years, the inquiry into human language from a complex network approach has arisen as a highly productive area, characterized by the convergence of disciplines such as statistical physics, systems science, linguistics, cognitive science, and natural language processing. This interdisciplinary endeavor contributes both methodological and theoretical insights into the idea of human language as a system. As we know, the idea of the language as a system traces back to the birth of modern linguistics and found its first extensive formulation in the work of Ferdinand De Saussure. The extent to which modern definition of the linguistic system and more or less overt relational definitions of meaning are consistent with the Saussurian original position is likely to become a matter of intense debate among linguists; the point we want to make here is that integrating the Saussurian intuitions on the structure of language taken from a linguistic point of view with what we now know about the functioning and organization of language at the mental level could lead to useful and interesting developments in research paradigms and cooperate in intensifying the efforts for an interdisciplinary approach to the structure of language.

4.5. Partial conclusions to our journey and open questions

A starting point from which to summarize our discussion is the diversity of disciplines and approaches involved in the debate. We started out saying that we would have been dealing mostly with psycholinguistic approaches, considering that the very need of studying language in the mind is, for the most part, a consequence of the cognitive revolution(s) in linguistics. The discussion we engaged in took us well beyond psycholinguistic facts, although at the same time it always brought us back to them. In fact, in order to explain linguistic behaviors and psycholinguistic data (such as slips of the tongue, the tip-of-the-tongue effect, different kinds of aphasia, and so on) many models of language comprehension and production have been developed, each of which highlighted a different aspect of what we could call our linguistic performances. Such models involved various kinds of assumptions, from linguistic to biological, to neurological, to cognitive ones, and elected a specific field of research as foundational for the analysis of the structure and organization of language in the mind.

In brief, serial models of language processing were based on generative and modularist linguistic and cognitive assumptions. In accordance to generative linguistics, meaning was understood as a dictionary in two ways: firstly, models were based on the possibility of distinguishing linguistic meaning from the rest of our knowledge and, thus, to analyze it in terms of its stable components. The possibility of decomposing meaning was instantiated in two way: on the one side, the meaning of a word can be decomposed into its basic traits (semantic or phonological) and, on the other side and prior to that, language needs to (and has to) be decomposed in its various aspects (phonological, morphological, syntactic, semantic), that are then separately observed and analyzed. The second consequence of generative and modularist views of language is that meaning is organized in our mind in the form of a dictionary, i.e. a list (or multiple lists) of elements organized by means of a certain principle to accommodate the needs of language retrieval and memory storage. The link with a modularist view of our cognitive architecture is evident: different aspects of language are listed in different modules that are then, in various ways, able to pass information one to the other. Hence, models stemming from such theoretical standpoints view language processing as a serial model in which the output produced by a lower level module is made available as input for a module is the level right above (this is the case for language comprehension, but if we invert it we can assume the same flow for language

production). Inevitably, serial models of language processing are tied to a serial and "symbolic" process of communication and information processing typical of early formulations in computer sciences. For the reasons described above, and for many others we have omitted, serial models (at least of the kind discussed here) do not stand any chance in explaining language processing in our real, everyday communicative life, which are fundamentally characterized by the interdependency and the interactivity of the various aspects.

Different assumptions and different outcomes characterize network approaches. Network approaches are grounded on the idea that the starting point for a discussion of language in cognition needs to be the biological and neurological structure that underlies our cognitive (and hence linguistic) abilities. Models of language processing and of the structure of the lexicon that derive from such perspective have been of fundamental interest in exploring, by means of computer simulations, aspects of our linguistic abilities that were completely left out by studies in the generative paradigm, such as the mechanisms that may underlie language learning or language decline. Despite their reductionist vibe and the often not explicitly discussed relationship (or confusion) between theoretical models, representational formats, and ways of implementations, network models of language processing and of the mental lexicon were able to overcome the limitations imposed by the generative paradigm of research. Network models, from Quillian's initial proposal to distributional models, to recent semantic networks, offer a different theoretical perspective. Linguistic theories based on networks approaches seem to offer a way to encompass (and to some extent be grounded on) those aspect of meaning and of the study of language that seem fundamental to grasp the nature and the organization of meaning from both a linguistic as well as a psychological point of view. We are here referring to the contextual, both co-textual and socially and culturally contextual, aspects of meaning, the fundamental pragmatic nature of meaning together with the biological constrains and scaffolds of the speaking subject, which is always and inevitably at the same time nature and nurture.

Our last stop in the journey into the mental lexicon was the discussion of free word association, the messy conglomeration of ways in which words are connected in our minds. Aside from the various issues that free word associations face in the light of both theoretical and empirical research, they allow us to abandon models of representation of idealized competences of idealized speakers and to tackle onto the actual and concrete ways in which speakers seem to deal with the complexity of linguistic meaning. An example of the fruitful applicability of word associations

and their representation into semantic networks is, for example, the re-definition of the notion of similarity: in defining similarity as the overlap of the distribution of associations between compared words, we can step away from a yes-or-no type of judgment about the similarity of two words and towards a graded account, which seems a much more realistic way of looking at the closeness of the meaning of two or more words. Furthermore, the data from free association tasks reveal that a strictly dictionary view of linguistic meaning is unlikely to be able to explain the complexity and variety of different association types that can be identified and to provide an understanding of processes of word retrieval, semantic priming, and so on. Thus, we acquired a renewed awareness of the need to endorse and theoretically start from an encyclopedic view of meaning according to which various aspects of our knowledge contribute in defining and organizing our linguistic abilities. Free word associations show the multifaceted nature of meaning, which indeed depends on the web of semantic relations, but it is also grounded on our knowledge of the world in terms of *frames*, and, most probably, on much more.

Within this framework, which has provided us with some fundamental insights in the study of the mental lexicon, we will argue that something is still missing. We think that the reflection on and the integration of some aspects that are central to linguistic account of meaning, but that have been generally overlooked by psycholinguistic approaches, might shed light onto yet unexplored corners of the universe of theories of the mental lexicon.

Some preliminary notes are required: what follows is not a properly structured theoretical proposal, but rather merely a set of reflections inspired by our review of the state of the art of the research on the mental lexicon. In addition, the few points we are about to briefly outline are not thought of as a critique to previous work; quite the opposite, our intention is to maybe set some seeds for further reflections and integrations of the precious approaches currently available.

Let us start from the easiest among the aspects we want to consider: the aspiration towards a truly multidisciplinary approach to the study of the mental lexicon. Our understanding of the mental lexicon necessarily needs to be grounded on a variety of different factors that go from the understanding of the biological-neurological architecture of our brain, to theories of cognitive mechanisms, to linguistic observations. It seems clear that an object of study such as the mental lexicon implies a system of interdisciplinary expertise. Understanding how humans process language, if there is something like a mental storage of our knowledge of language, and how it is organized, necessarily need the contribution of: biology for what we know about how the brain

processes information and what areas or functions of the brain are involved in dealing with language, if there is anything specific about language or not; psychology and cognitive sciences because language is undoubtedly a cognitive phenomenon; linguistics because language is not just an individual cognitive fact, but it also is a intrinsically social and cultural fact that, as much as all the rest, contributes in shaping our linguistic abilities. This may seem a trivial point, but it is definitely easier said than done. For example, it is not always easy or straightforward to evaluate the cognitive plausibility of a theory of meaning and, if we start from linguistic theories, designing psycholinguist experiments on the basis of certain linguistic assumptions or intuitions is an insidious and slippery terrain. We would like to offer a practical example of an idea in linguistics about the nature of meaning that has not yet been taken up by the psycholinguistic research on the mental lexicon, at least to the extent of our knowledge, and that would most probably lead to interesting research. Underlying the majority of the approaches we have reviewed in this chapter there is the idea that the plane of the expression and the plane of the content of the linguistic sign can be analyzed separately and are, in some ways, organized according to different principles. As we know from our brief discussion of the Saussurian positions (see chapter 2), the very nature of the linguistic sign is to carve up, at the same time, the plane of the content and that of the expression, making the two inseparable. Understanding the linguistic sign as the inseparable combination of content and expression rules out any account of the mental lexicon and of lexical processing that assumes the strict separability of the two components. Some of the accounts discussed above seem to implicitly point towards this direction; for example, as we have seen when discussing the cohort model (*4.2.3)*, the detection of a word in a heard utterance is not independent from the semantic, syntactic, and contextual constrains, suggesting that both the level of the expression and the level of the content play an intertwined role in both the detection and interpretation of the linguistic input.

This brings us to our second point: the need to reinstate a dynamic perspective in the discussion of the mental lexicon. The processing of language should not be viewed as a process of coding and decoding of linguistic elements, but rather as a constitutively interpretative process. As De Mauro variously pointed out in his interpretation of Saussure's thought, it is by means of the semiotic activity, which instantiates and uses the linguistic codes, that we connect the system of the classes of the plane of the expression to the system of classes on the place of the content. Even more specifically to the point we are discussing here, De Mauro (1994) argues that in order to

understand the process of linguistic comprehension we need to start from the concrete act of *parole* and we ought to assume a semiotic and psycho-biological perspective.

> [N]elle lingue la determinazione dell'appropriato valore semantico di una forma è data dalla messa in rapporto tra dato formale, segnico, e dati sostanziali e informali. […] [C]iò comporta che i dati sostanziali e informali possano essere un insieme esplorabile di eventi, entro i quali emergono regolarità quando si osservi come essi si compongono e ordinano in rapporto al riconoscimento della collocazione della produzione linguistica particolare nella *massa parlante* e nel tempo storico, secondo quanto ha insegnato Saussure (in generale inascoltato su questo punto), e in rapporto, quindi, alle presuntive peculiarità del locutore, della sua persona, della sua *parole*. Le peculiarità possono ben ridursi al minimo, mascherarsi, come avviene specialmente nella produzione scritta, dietro formulazioni standard della *parole*. E tuttavia anche tali formulazioni standard sono pur sempre solo una delle possibili opzioni, e sono dunque esse stesse a loro modo tracce peculiari di individualità in un dato momento, di un *hic et nunc onmimodo determinatum* (De Mauro, 1994, p. 57).

Hence, the study of how we process language is inherently a study of the *parole*, of the performance, and not of the idealized competence of the speaker. Our definition of the processes of comprehension, and of the supposed storage we should have to make the process work, need to be able to account for all that in language is *unexpected*, as in fact each instantiation of language, every act of *parole*, essentially is. Each and every act of language cannot be de-codified *per se*, but only within a broader interpretative perspective.

According to this perspective, the notion of the mental lexicon could be reinterpreted in the sense that the aspect we need to account for is not necessarily how we recover stored, pre-packed knowledge from memory, but instead how we construct that knowledge each time we use language, each time we *do* meaning (cf. § 2.3.4). Hence, meaning is not necessarily something that *is there*, but rather something we create each time we are engaged in a semiotic act. This does not imply that no knowledge needs to be stored. It simply adds a piece to the puzzle of the various aspects that should be taken into consideration when elaborating theories of language processing and of the mental lexicon. In particular, it seems important to highlight that understanding language processing as a mere coding and decoding of inputs and output happening only within our minds is a reductive approach to language that overlooks important aspects of it, such as the fact that our "shared" language – language as a cultural phenomenon - and our linguistic environment - what we have previously called the language outside the mind - play an important role in shaping our (internal) knowledge of language. In the next chapter, we will come back to

this issue and we will try to show where the integration of a non-strictly internalist view of language and of meaning might lead us. We will begin again with discussing the cognitivist idea that language equals with concepts, and more precisely with conceptualization, and we will show how seeds internal to the cognitivist perspective can take us to highlight in what sense the use of language is foundational in our construction of meaning.

In the first chapter, we already introduced the notion of *words-as-cues* (§ 1.2.3). The idea that words do not have meaning, but that rather are cues to meaning (i.e. they activate mental states), allows us to start delineating a somewhat different perspective inasmuch as it opens up to the possibility of identifying a mind-external level of language. Elman's proposal of a lexical knowledge without a lexicon (see § 4.3.4), offers a first insight on how the properties of our linguistic environment might be relevant in shaping the content of our knowledge. But further reflections and investigations need to be conducted in order to discuss a possible overlap of the two ideas which, indeed, start from two very different theoretical standpoints and developed considerably different lines of research.

One aspect of neural networks that we have only briefly mentioned earlier, but that is a fundamental connection with the following chapter and the direct consequence of the line of thought we have been developing up to now, is that, in order to learn, neural networks are fed with inputs and outputs from natural language or from simplified versions of natural language, often from large linguistic corpora. This means that most neural networks learn from language; i.e. that language (as a fact external to the mind) is fundamentally involved in the way in which we build our internal representation of language. Clearly, this does not mean that neutral networks learn what they learn in the same way we learn what we learn; but both models and humans seem to learn from language itself. What do networks learn from language and what do they allow us to infer about the structure of language? In the most recent years, the intersection of neural networks implementations and of distributional theories of linguistic meaning has given rise to a whole new set of theories and models that allow us to address and observe the structure and organization of the language system starting from linguistic data. This will be the focus of the next chapter.

Chapter 5

Finding structure in uses: the distributional semantic approach

---

5.1. Setting the path towards a "language-specific" level of linguistic meaning

In the previous chapter, we took a long side road into the notion of the mental lexicon that allows us to uncover some of the fundamental characteristics of psycholinguistic approaches to the structure of language, and more specifically of the lexicon, for the purposes of language comprehension and production. One general conclusion, which applies to different extent to the different approaches we discussed, is that the structure of language is seen as something that is strictly internal to our minds and/or as a product of the way in which language is internally processed by our cognitive system(s), whether it would be by means of a "boxes and arrows" or a "nodes and weights" model. Indeed, based on the kind of model we theoretically embrace, we end up with very different cognitive architectures for language processing and with opposite kinds of lexical storage: our language system comprises either a set of syntactical and compositional rules and a symbolic and fixed set of linguistic-conceptual representation, or a dynamic web of word-concepts associations or of distributed representations. What comes out as a striking evidence from the framework outlined in the previous chapters is that, if we analyze language and language structure from a psychological and psycholinguistic perspective, there seems to be nothing specific, i.e. nothing specifically linguistic, about semantic processing or semantic representations. Hence, the idea that words essentially label (configurations of) concepts remains one of the main tenants of cognitive lexical semantic approaches (Riemer, 2015).

This comes as no surprise given that saying that something about linguistic meaning may be (even partially) "specifically linguistic" immediately flashes some red lights warning us that we are stepping onto a minefield. As we have extensively discussed in *Chapter 3*, one of the fundamental achievements of the second cognitive revolution (which gave rise to a fundamentally

new theoretical and practical disciple generally address as cognitive linguistics[123]) against formalist, structuralist, and generative paradigms was the non-autonomy, hence the overall non-linguistic-specificity, of linguistic meaning. Cognitive linguistic approaches to lexical meaning emerged as a rejection to the generativist view of language (and specifically of grammatical knowledge) as separated from the rest of our cognitive abilities and processes. Conversely, according to cognitive linguistics, language cannot be studied without reference to more general principles of human cognition and to the socio-cognitive framework in which lexical meanings are inextricably associated with language users' bodily, perceptual, and cognitive experiences of cultural and historical phenomena. The meaning we associate with words is constrained as well as promoted by subjective and intersubjective conditions, thus nullifying any strict dichotomy between linguistic and encyclopedic meanings. Lexical networks of the kinds discussed in the previous chapter (we are especially referring to local representations networks and networks based on free association data) are the clearest exemplification of the encyclopedic nature of meaning representations. Knowing what a *sleepover party* is not only requires the juxtaposition of *sleepover* and *party* together with the possibility of composing the meaning of *sleepover* as the sum of the meaning of *sleep* and the meaning of *over* in the appropriate construction, but also relates to, for example, acceptable social behaviors or other kinds of pragmatic aspects that themselves are part of the meaning we associate with the expression *sleepover party*. Similarly, knowing the meaning of *dog* implies the reactivation of a *frame* (or maybe even multiple *frames*) in which our knowledge of what a dog is, what it needs, the activities we can do with it, and so on, is organized. The conceptual structure (i.e. the conceptual network) that provides the conceptual content to linguistic expressions can range from fairly simple concepts or a perceptual experience to complex knowledge clusters (for similar examples and discussions cf. Lemmens, 2015). Our conceptual structure, and consequently (in cognitivist approaches) the semantic content of language, is grounded in our cultural and physical experience. Hence, cognitive semantics essentially emerged as a 'maximalist' form of semantics, in which the structuralist heritage of isolating semantics within language is substituted with a contextualized approach, in which natural language semantics is seen against the background of different types of context: human psychology, language use, and a broader cultural and historical canvas.

[123] We have discussed the difference between first-generation cognitive linguistics and second-generation cognitive linguistics in chapter 3 (§ 3.4.).

Especially after the last twenty years of research in psychology and in the cognitive sciences, there is no room for doubts that our concepts and the meaning of words are grounded in perceptual and cultural experience and that language, as well as our cognitive structure and contents, should be always considered against such experiential and contextual background. What seems more controversial and theoretically loaded is the often-implicit consequence that arises from such cognitive approaches to language: linguistic meaning is not just grounded in our bodily, psychological, and cultural experience, but, at the same time, referring back to either one or a combination of such broadly-understood contextual factors seems to be enough to fully comprehend and define the nature and structure of linguistic meaning. In very general terms, the study of the meaning of words dissolves into the study of these contextual components of meaning and no specific place is reserved to a "language-specific" level of meaning. Let us take one step at a time and consider the two main tenants of cognitive linguistic approaches: the cognitive perspective on meaning (which has been the focus of Chapter 3 and thus will be only briefly analyzed here under a different light) and the usage-based perspective on language (on which we will more specifically focus at this stage of our argument as it leads the way to possible interesting developments in the study of lexical semantics). As for the cognitive perspective on meaning, much has already been said throughout our work (§§ 2.1.2; 2.1.3; 3.4.1; 3.5.1) so we will limit our discussion here to some final remarks. A little more space will be given to the notion of meaning as conceptualization, which we have already mentioned before but we have not specifically addressed.

5.1.1. Last remarks on meaning as concepts

According to cognitive linguistic approaches, word meanings are psychological entities located in people's minds and, as such, straightforwardly map onto the conceptual structure.[124] A conceptualist account of meaning remains a central argument of both the first and the second cognitive revolution, between which the cut-off line is the format and nature of the conceptual representations, and consequently of meaning representations, but not their cognitive nature (apart

[124] In Chapter 2 and 3, we have seen how there actually are many levels at which such identification of linguistic categories and conceptual categories has been theorized and dealt with, and that not all accounts equally share the same strength in endorsing such view on meaning. We will come back to this point in more detail shortly.

from this commonality in understanding linguistic meaning as conceptual, generativist and second-generation cognitive linguistic accounts differ in all essentials; for a discussion cf. Paradis, 2003, 2012).

Linguistic meaning is indeed conceptual in some weak sense, i.e. in the sense that it is an internal mental factor supported by some sort of psychological structure, avoiding in this sense any strong commitment about the nature of the structure in question. Nonetheless, as Riemer (2015) points out, it is not on this weak notion of conceptual meaning that the theoretical dispute has taken and still takes place, but on the much stronger claim that the structure and content of word meaning can actually be fully identified with the conceptual structure (of whatever format we might imagine it), which in turn leads to a more or less explicit and overarching identification of linguistic categories with conceptual categories, both in their content and in their structure. As a result of the full conceptualization of meaning, studying language and its structure offers a window on the organization of our cognitive content, of which language is thought of as a reliable mirror.

Throughout our discussion in the previous chapters, we have tried to point out the shortcomings of such identification of language and concepts; narrowing it to the bone, our main concern is that embracing a non-language-specific view of linguistic meaning leads to the point of favoring a mere communicative function of language and of underestimating, if not completely abandoning, the interest for a possible cognitive function of language, which is bound for being inconsistent if linguistic meaning simply overlaps with conceptual content. Not to mention that the question of how lexical semantic explanations can be related to psychologically plausible models of mental representations does not have a straightforward commonly accepted answer, and has acquired a new urgency in light of the fact that many traditional accounts of word meaning in terms of veridical statements about the world are not supported by recent research in the cognitive sciences (for an important discussion on the topic cf. Riemer, 2013, 2015), and that many psychological accounts of word meanings (and of representations in general) no longer support the classic notion of modality-independent, discrete, stable concepts. Hence, in order to contribute to a theory of the mental lexicon, lexical semantics can no longer rely on some assumed psychological notion of ‘conceptual content’ that is, in turn, thought of as an exhaustive definition of meaning.

Purely conceptualist accounts of meaning[125] seem to fall short on at least two fundamental aspects. Let us start from an example: dual-stage, strictly modular models of linguistic processing had soon to give way to dual-stage dynamic models that, even within a modular perspective, recognized the need to account for the interconnectivity of the language modules in explaining language production and comprehension. This example, among others, shows how, on the one side, our understanding of linguistic meaning changes according to newly acquired knowledge (or even merely on the basis of suppositions) about our conceptual structure, and how, on the other side, the opposite may be true as well: cognitive theories of language can be partially shaped by the need of accounting for mechanisms and structures observable in language use. This observation leads us to the second shortcoming of purely conceptualist account of meaning; if we assume that linguistic content is equal to conceptual content, then the simple observation that languages differ in the way they cut up a range of perceptual experiences, i.e. in the way languages delineate boundaries in the conceptual content, leads to the unfortunate conclusion that the same set of universal cognitive operations give rise to different conceptual worlds that are then portrayed in language. Hence, if we assume the perspective of first-revolution cognitivists, we can easily see why meaning had to be thought of as a mere surface phenomenon, whereas the core and structure of language was reserved to a universal set of principles that constituted the fundamental grammar shared by all languages.

Initiators of the second cognitive revolution took a very different stand to this regard. Although essentially endorsing a conceptualist view of meaning, the very definition of meaning and of what accounts as meaningful in the domain of language profoundly changed, challenging most of the major assumptions of generative grammar in the way it understood both language and cognition more broadly. As a consequence of the distinction between syntax and semantics endorsed by Chomsky, in the immediate post-Chosmkian era, psychologists interested in meaning have mostly focused their attention on problems other than syntax. Psycholinguists have focused on such things as lexical access, discourse processing, and speech perception, whereas developmental psychologists have focused on such things as lexical development, pragmatics, and adult-child conversation. The novelties brought by these approaches in the study of meaning are groundbreaking: cognitive linguistics spelled out a renewed view of the notion of category, that

[125] Fodor's *mentalese* is just the most famous example of extreme stands about the conceptual nature of meaning (e.g. J. A. Fodor, 1975).

from discrete and bounded became gradient and flexible in their boundaries, and grounded our concrete as well as more abstract knowledge not in the manipulation of abstract (often innate) amodal symbols processed independently from the rest of our cognition, but rather in modal representations grounded in our bodily experience of the world, dependent on and integrated with domain-general cognitive functions and operations. However, despite the revolutionary insights that were provided in the study of cognition, not much changed in the view of the role played by language. Words essentially map onto our cognitive representations of the world, leaving aside any possible specificity of linguistic meaning.

5.1.2. Meaning as conceptualization

In recent years, a similar shift in perspective towards a more semantically oriented consideration of meaning occurred in the study of grammar and was characterized by the neat divergence from the mathematical framework of the generative approach towards a cognitive-functional view of language (Tomasello, 1998a). A new family of approaches to language structure, often growing out of the Chomskian paradigm as several of its practitioners became more and more dissatisfied with its narrow scope and its neglect of the cognitive and social dimension of language, encompasses a variety of approaches that differ from one another in a number of ways but that are united in the rejection of the autonomy of syntax position. In diametric opposition to the theoretical and methodological assumptions of the Chomskian paradigm, cognitive-functional linguists take as their object of study all aspects of natural language understanding and use, including unruly idioms and irregularities. They take as an important part of their data not disembodied sentences derived by introspection, but rather utterances or other longer sequences from naturally occurring discourse.

On such basis, cognitive-functional approaches to language structure developed an entirely new paradigm of study within the field of the cognitive sciences, giving rise to various paradigms of research from Langacker's *Cognitive Grammar* (1987, 2008) to a variety of constructionist approaches (e.g. Bergen & Chang, 2013, 2005; Croft, 2001; Fillmore, Kay, & O'Connor, 1988; A. E. Goldberg, 1992, 1995, 2006). Cognitive Grammar and constructivist approaches do not completely overlap in their theoretical or methodological stands (similarly, we observe a variety of different positions within constructivist approaches). Nonetheless, for the purposes of our

argument, we will treat them as a more or less unitary phenomenon as we are interested in outlining some of the core aspects, essentially shared by all mentioned approaches, that constitute a part of the (evolution of the) historical and theoretical background on which we are setting our proposal, especially in regards to the renewed formulation of the nature and function of language in cognition.[126] We will focus our attention on two main tenants of constructionist approaches and their implications for the study of meaning: the *semiological function* of language (e.g. Langacker, 1998), exemplified by the notion of *symbolic structure* or *construction,* and the usage-based perspective, which is a foundational aspect of these approaches.[127]

According to cognitive linguistics, language is shaped and constrained by the functions it serves. "Language has two main functions: a *semiological function,* allowing thoughts to be symbolized by means of sounds, gestures, or writing, as well as an *interactive function*, embracing communication, expressiveness, manipulation, and social communion" (Langacker, 1998). In contrast to formalist traditions (such as generative grammar), the recognition of the foundational status of these functions is the primary distinguishing feature of functionalist approaches. Cognitive linguistics, and especially cognitive approaches to grammar, stand out by emphasizing the semiological function of language and the crucial role of conceptualization in social interaction. The semiological function is instantiated in the formation of *symbolic structures* at all levels, from morphological entities to whole sentences:

> [the] central claim is that grammar is per se a *symbolic* phenomenon, consisting of patterns for imposing and symbolizing particular schemes of conceptual structuring. It is held that lexicon, morphology, and syntax form a continuum fully describable as

[126] Henceforward, for the sake of simplicity, we will address all approaches as constructionist approaches (to the language structure), if not otherwise indicated. For a discussion and review of the different approaches cf. *The Oxford Handbook of Construction Grammar* (2013b), with specific reference to the useful introduction of Hoffman and Trousdale (2013a), and the volume *The New Psychology of Language* edited by Tomasello (1998b). For a discussion of the connection between *Cognitive Grammar* and constructivist approaches see Broccias (2013).

[127] In order to make the exposition accessible and coherent and to point out the relevant aspects that are of interest to us, we will separately discuss the two tenants although they are interrelated and co-foundational of constructivist approaches. A third fundamental pillar of cognitive approaches to grammar is that language is grounded and emerges from language-independent cognitive processes that concern the categorization and conceptualization of experience (in a similar fashion to the positions endorsed by second generation cognitive linguists interested in the structure and origin of categories we have discussed in Chapter 3). Language is an integral part of human cognition and any account of linguistic structure should therefore articulate with what is known about cognitive processes in general. We will not address this third pillar extensively in our discussion as our main interest is in outlining the notion of meaning that underlies cognitive approaches to grammar (Langacker, 1987; for a recent discussion cf. Bybee, 2010). Notwithstanding, the idea that language is grounded in non-linguistic psychological processes is fundamental in order to understand the nature of meaning; semantic structure is conceptualization tailored to the specifications of linguistic conventions. As we will see, meaning is conceptualization and it arises from the operations carried forward by non-language specific cognitive processes.

> assemblies of *symbolic structures* (form-meaning pairings), and consequently, that all valid grammatical constructs have conceptual import. Reducing grammar to symbolic assemblies affords a major theoretical unification as well as great austerity in the kinds of entities analysts are allowed to posit. It is realistic from the psychological standpoint, for unlike semantics and phonology grounded in the broader realms of conceptualization and sound grammar is not connected to any independently accessible domain of experience. It is more naturally taken as residing in schematized representations of sound-meaning pairings, abstracted from (and immanent in) the specific symbolic configurations observable in complex expressions. Described in this manner, grammar (like lexicon) can be seen as directly reflecting the semiological function of language (Langacker, 1998, p. 2).

In summary, cognitive grammar (as well as constructionist approaches) explore the idea that form-meaning pairings might not only be a useful concept for describing words and morphemes but that perhaps all levels of grammatical description involve such conventionalized form-meaning pairings (i.e. construction). All constructions (e.g. morphemes, words, complex words, idioms, passive construction, and so on) are part of a lexicon-syntax continuum, inasmuch as grammar itself is meaningful and meaning resides at all levels: "grammar is simply the structuring and symbolization of semantic content" (Langacker, 1987, p. 12). In order to understand in what sense grammar, as well as morphemes or lexemes, are meaningful we need to directly address Langacker's notion of *symbolic structure* and of meaning as conceptualization, which are the fundamental tenants of cognitive grammar and (even if sometimes not overtely) of constructionist approaches to grammar.

Symbolic structures (hence constructions) are form-meaning pairings: they consist of a *semantic pole,* a *phonological pole,* and the association between them (see figure 17). The semantic space (or pole) and the phonological space (or pole) can be seen as parts of the human cognitive organization. In particular, the semantic space can be thought of as "the multifaceted field of conceptual potential within which thought and conceptualization unfold; a semantic structure can then be characterized as a location or a configuration in the semantic space" (Langacker, 1987, p. 76). Similarly, the phonological space is our range of phonic potential. By coordinating the two, we obtain the symbolic space and symbolic structures, which can be characterized as a configuration in symbolic space: "a symbolic structure consists of a semantic structure at one pole, a phonological structure at the other pole, and a **correspondence** linking them together" (Langacker, 1987, p. 77 - bold in text). Morphemes are the minimal symbolic units and the building blocks of grammar, which consists in the successive combination of symbolic structures to form

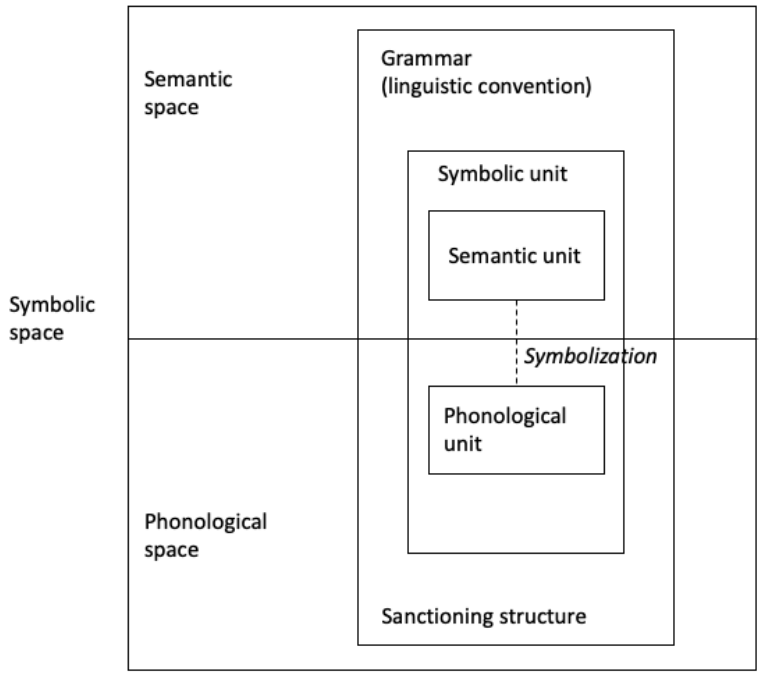


Figure 17. The notion of symbolic unit (adapted from Langacker, 1987, p. 77)

progressively larger symbolic expressions. Hence, grammar is referred in the broadest sense to the system of linguistic knowledge in the mind of the speaker (for a discussion cf. for example V. Evans & Green, 2006).

It has been variously pointed out how the cognitive notion of symbolic structure has its roots in the Saussurean notion of linguistic sign (for the foundational definition cf. Langacker, 1987, pp. 11–12; for a review cf. Hoffmann & Trousdale, 2013a). However, Langacker's definition of symbolic structure, as well as subsequent definitions of constructions, although drawing from the Saussurian notion of sign as the relation between a conceptual plane and a phonological plane, substantially leave aside the notion of *arbitrariness,* which is central in Saussure's view, favoring conventionality instead. In more details, the semantic aspect of a symbolic structure is not seen as the resultant of the operation of symbolization, but rather as a prerequisite of it. In this sense, meaning (which is understood as conceptualization) pre-exists the sign and it is by means of the process of *symbolization* (Langacker, 1987, p. 77) that the correspondence between the already-defined structure of the semantic space and the phonological space is established. These form-meaning pairings become conventionalized by means of language use in social interactions.

Hence, the notion of symbolic structure, although drawing from some Saussurian insights, it is not of Saussurian descent inasmuch as it almost completely neglects the notion and the role of *arbitrariness* and, consequently, the relational view of linguistic meaning.

In fact, Langacker's notion of symbolic structures rests on a very elaborate conceptualist view of meaning.[128] An expression's meaning involves content and construal, i.e. the conceptual content it evokes and how that content is construed. Meaning (in the sense of content) derives from embodied human experience and is to be found in the minds of speakers who produce and understand the expressions. In the early stages of cognitive semantic research, linguistic structure is immediately based on conceptual structure; nonetheless, it soon became clear that a one-to-one correspondence of words and concepts could hardly be endorsed and, although there is reasonable evidence that language structure and cognitive structure are related, this relationship seems not to be of a direct kind. Hence, the focus shifted from concepts to conceptualization, from a static to a dynamic view of meaning, and to the role of usage and development in the shaping of it. For the purposes of studying language as a part of cognition,

> [t]he conceptualizations we entertain are undeniably internal, in the sense of taking place in the brain, yet reach beyond it in the sense of being conceptualizations of some facet of the world. In speaking, we conceptualize not only what we are talking about but also the context in all its dimensions, including our assessment of the knowledge and intentions of our interlocutor. Rather than being insular, therefore, conceptualization should be seen as a primary means of engaging the world (Langacker, 2008, pp. 28–29).

Hence, meaning is always dynamic and at the same time internal to the individual speaker and shaped by the interactions at the social level. Being identified with conceptualization, i.e. with the possible various ways in which we engage with the world by means of cognitive processing, meaning encompasses any facet of mental experience. In line with most perspectives in cognitive semantics, semantics is encyclopedic in nature and lexical meaning resides in a particular way of accessing an open-ended body of knowledge pertaining to a certain type of entity. The elaborate

[128] Unlike other constructivist approaches, the notion of meaning is particularly complex in Langacker's Cognitive Grammar. Here, we will only discuss this conceptualist notion of meaning in very general terms trying to outline its main characteristic (for an in-depth discussion cf. Broccias, 2013). One important aspect to point out about Langacker's approach to grammar is the rejection of a truth-conditional view of meaning which led to the autonomy of grammar position. In contrast, Langacker argues that "the meaningfulness of grammatical elements becomes apparent only by adopting a conceptualist semantics that properly accommodates our ability to conceive and portray the same objective situation in alternative ways" (Langacker, 1998, p. 3).

conceptualization that takes place when we understand or produce a sentence is only prompted by lexical meaning and compositional patterns: “an expression’s meaning includes as much additional structure as is needed to render the conceptualization coherent and reflect what speakers would naively regard as being meant and said, while excluding factors that are indisputably pragmatic and not necessary to make sense of what is linguistically encoded” (Langacker, 2008, p. 42). Thus, semantics and pragmatics do not collapse one onto the other, but rather form a gradation with no precise boundary between the two.

If, as we have seen, content is essentially conceptualization; what about *construal*? *Construal* is fundamental to cognitive semantics as it allows to theoretically and empirically tackle onto the fact that we can conceptualize the same reality in many different ways. Hence, meaning is a function of the conceptual content and of how that content is construed.

> As an inherent aspect of their conventional semantic value, linguistic elements impose a particular construal on the content they evoke, and speakers adopt it for purposes of linguistic expression. Differences in conventional patterns of construal are largely responsible for the impression that languages embody contrasting "worldviews." A strong claim of linguistic relativity, that language drastically and unavoidably shapes thought, is not however warranted. The effect of linguistically imposed construal may be fairly superficial (Slobin, 1987). The content evoked by expressions can be largely the same (affording the basis for understanding and rough translation) even when construed very differently (Langacker, 1998, p. 4).

As part of its conventional semantic value, every symbolic structure portrays (construes) its content in a certain fashion. Theoretical constructs (ways of organizing the conceptual content) are, for example, notions like *domain, profile, trajectory, vantage point, scanning, mental space, immediate scope*, and so on, on the basis of the fact that they are in line with what we know about non-linguistic cognition (e.g. about vision) and that they allow to provide an adequate semantic description of many linguistic phenomena (Langacker, 2008).

As we can see from this very brief discussion of the notion of construal, although pointing out that languages differ in the way they conventionalize certain constructs, the mechanisms by which such constructs come into being are in line with the fundamental tenant of cognitive semantics: language is essentially shaped by non-language-specific cognitive mechanisms. Constructionist approaches often turn to language-external explanations, such as universal functional pressures, iconic principles, and processing and learning constraints to explain the mechanisms by which languages conventionalize form-meaning pairing as well as cross-linguistic

generalizations (for references cf. A. E. Goldberg, 2003). Hence, the content evoked by a certain construction can be largely the same, given the complex process of internal meaning construction, even in those cases in which the way in which it is construed is very different.

5.1.3. Usage-based approaches to grammar

As we have seen, meaning is internal to the mind as conceptualization is a matter of cognitive processing, and, at the same time, it is always negotiated in social interactions. The structure of language is not a fixed set of core rules shared by each and every language, but rather it emerges from patterns and standardized repetitions in language use: "grammar is a dynamic system of emergent categories and flexible constraints that are always changing under the influence of domain-general cognitive processes involved in language use" (Diessel, 2015, p. 256). The mental grammar of speakers is shaped by use and development, i.e. by the repeated exposure to specific utterances in which domain-general cognitive processes, such as categorization and cross modal associations, play a crucial role in the entrenchment of constructions.

The basic premise of usage-based theories is that experience with language creates and impacts the cognitive representation for language. The term was officially coined by Langacker (1987, see also 1990 chap. 10) but it is a principle shared by the majority of cognitive-functional approaches, which argue that grammar is created by the conventionalization of commonly used and often repeated discourse patterns (Bybee & Beckner, 2015). Even more radically, recent proposals in usage-based theory argue that grammar is "one's experience with language" (Bybee, 2006, 2010): aspects of that experience, for example the frequency of use of certain constructions or particular instances of constructions, have an impact on mental representations that is evidenced in speaker's knowledge of conventionalized phrases and in language variation and change.

Two further fundamental aspects of usage-based theories are: the centrality of language development - how language is learnt and constrained by domain general processes such as categorization, rational inferences, and social cognition (Goldberg, 1995, 2006) - and the theoretical and methodological advancements in episodic memory theory (Bybee, 2006, 2010, 2013).

Episodic models propose that memory for linguistic experience is like memory for other types of experience: each token of experienced linguistic behavior has an impact on cognitive

representations. When stored representations are accessed, the representations themselves change. In addition, memory storage for linguistic experience includes detailed information about the tokens that have been processed, including the form and contexts in which they were used. Exemplar representations contrast with the more abstract representations of structural or generative theories (at all levels—phonetic, morphosyntactic, and semantic/pragmatic) and allow for specific meaning from context of use to impact cognitive representations both in the development of specific meanings and implications as well as in developments of specific constructions.[129]

In general, usage-based theories stand in sharp contrast to structuralist and generative approaches, which hold that the study of the linguistic system has to be kept separated from the study of language use (*langue* vs. *parole*[130] and *competence* vs. *performance*).

In diametrical opposition to the methodological assumptions of generative approaches, cognitive-functional approaches, hence construction grammar, take as their object of study all aspects of natural language understanding and use, and as an important part of their data not disembodied sentences derived from introspection, but rather utterances or other longer sequences from naturally occurring discourse (Tomasello, 1998a). The aim of cognitive-functional approaches is not to discover a universal set of constructions or compositions, but rather to detail the "structured inventory of symbolic units" that make up particular natural languages (Langacker, 1987). Such structured inventory can be thought of as a dynamic system of emergent structured and flexible probabilistic constraints shaped by communication, memory, and processing.

Challenging the widespread assumption that linguistic structure is built from a predefined set of innate linguistic concepts, usage-based linguists conceive of language as a learnt dynamic network in which the various aspects of a language user's linguistic knowledge are constantly restructured and reorganized under the continuous pressure of performance. Hence, language is neither self-contained nor well-defined and metaphors that allude to a conception of language as a

[129] There is a growing body of literature in psycholinguistics showing evidence for episodic memory theories (e.g. Bybee & Hopper, 2001; Goldinger, 1998; Pierrehumbert, 2001, 2003). Models of this kind have been very successful in a wide range of applications in cognitive psychology, and have recently become very popular in addressing word-specific variation in fine phonetic detail, context-sensitivity of such details, and gradual language change (e.g. Pierrehumbert, 2001). As we will see in the following sessions, episodic models are in line with the stronger, cognitive version of the distributional hypothesis (e.g. § 6.1.1).

[130] Although such an understanding of a strict distinction between *langue* and *parole*, language system and use, is generally acceptable for post-Saussurian structuralist approaches, the relation between *langue* and *parole* in De Saussure's view is far more articulated and complex. We have already reviewed and discussed this aspect of the Saussurian perspective in Chapter 2 (§§ 2.2.3; 2.3.3.) and we will briefly come back to it at the conclusion of the paragraph.

distinct, discretely bounded, clearly delimited entity, and to a language system as a stable and separate mental component, fail to recognize and characterize language as a dynamic "socioculturally grounded cognitive activity" (Langacker, 2008, p. 216).

Consequently, usage-based construction grammar approaches often rely on the metaphor of language as a network of relationships to understand, represent, and explore the enormous inventory of conventional linguistic units abstracted from usage events.[131] For example, connectionist models (of the kind discussed in Chapter 4) allow to represent language knowledge as stored not in the form of items or rules, but rather in the form of changes to the strengths of connections among simple processing units. Distributed connectionist models, like exemplar models, exhibit sensitivity to both general and specific information. They allow to effectively derive representations through the processing itself, and to capture the structure present in the set of items these networks are tasked to learn. For instance, SRN is able to discover the relevant linguistic features by assigning similar representations to words with similar sequential constrains of their co-occurrences with other words in the sequence. In conclusion, both learning and processing within such models is inherently gradient in nature and thus inherently compatible with the gradience and gradual change we see everywhere in language. The learned internal representation of an item can overlap with, and hence share in the knowledge the network has about, other related items to a varying degree. Furthermore, if distributed information in the input changes a little, the network can adjust its weights, so that there is a gradual change in representation corresponding to the shift in the distribution. These properties contribute to our understanding of the processing underlying human knowledge and use of language and are in line with a dynamic view of the structure of language as described in constructionist approaches to grammar (Bybee & McClelland, 2005; McClelland & Bybee, 2007).

We have thus come to outline a first way in which we find structure in use: usage-based construction grammar approaches focus on the way in which we abstract language knowledge (networks of constructions, i.e. form-meaning pairings) from language use by means of general non-language specific cognitive processes, such as, for example, categorization and chunking. Our

[131] In the literature, there are various examples of usage-based construction grammar network representations each of which highlights different kinds of structures and of relations between the constitutive elements (e.g. Bybee, 1985; A. E. Goldberg, 1995; Langacker, 1987, 2008). Since we are not offering a full review of construction grammar approaches nor we aim at evaluating them, but rather we are interested in outlining some of the general aspects for the purposes of our discussion, we will not discuss the different network models in detail (for a review cf. for example Diessel, 2017).

aim is to make a first step towards arguing that a similar approach can be taken in the study of linguistic meaning. The status and advancement of our research does not allow us to make strong statements about the cognitive plausibility or the strength of a usage-based account of meaning, nor to propose a full-fledge theory. We rather aim at setting the path and opening the way to a wide variety of possible new lines of inquiry in the study of language from a psychological and psycholinguistic perspective. More specifically, we will argue that a usage-based approach to meaning is in line with the claim of a "language-specific", relational and differential (*a la* De Saussure) level of linguistic meaning, and that arguing for a "language-specific" level of meaning is not in contrast, but rather in line, with cognitive approaches to language.

5.1.4. The "language-specific" level of linguistic meaning

In this and the next chapter, we will try to outline what we mean by "language specific" level of linguistic meaning and we will argue that such a level is to be found in the systematicity of language usages and is observable by means of the distributional semantic approach.

While outlining our actual proposal, it seems fundamental to clearly state what our argument is not about. The proposal of identifying a "language-specific" level of linguistic meaning is not a relapse in the autonomous view of meaning and it does not set itself in contrast to the encyclopedic and internalist views. For instance, we endorse the cognitive notion of meaning as conceptualization and of words (or constructions in general) as cues to large body of encyclopedic knowledge if taken from the perspective of the cognitive processing of the individual speakers, while underlying that the way in which languages organize and segment the conceptual space (i.e. they way in which languages organize meaning) is a product of the standardization of repeated language usages, while indeed being bounded by our perceptual and cognitive apparatus.

In this sense, we are drawing from the Saussurian notion of sign and of radical arbitrariness of the signs; in contrast to, for example, Langacker's notion of sign, in which meaning as conceptualization is pre-given to the formation of the sign, which in turn is a matter of conventionalized communicative practices, we argue that the linguistic meaning of words (or larger constructions) is established in the very formation of the linguistic sign as a dynamic unit that is always in relation to other units in the system and that takes its value from these relations as much as from the non-linguistic organization of our mental content.

Linguistic signs are not just a matter of conventionalized form-meanings pairings, in which a form is attached to a meaning (whether would it be a concept or a process of conceptualization). Linguistic signs carve up nature at it joints as much as they reflect the joints of nature, according to the principle of radical arbitrariness.[132] We can look back at Langacker's notion of construal and add the relational nature of the language system to it, while maintaining its fundamental psychological nature. In this sense, the ways in which languages organize meaning, i.e. the way in which languages portray content in a certain fashion, is a matter of not only psychological mechanisms, but, at the same time, of the language-specific relational pertinentization of content: linguistic units obtain (part of) their meaning from the relations they entail with other elements of the language systems, delineating boundaries of pertinencies that are not entirely explainable and justifiable by conventionality and by the operations carried on by cognitive general mechanisms.

The process of "doing meaning", which can be thought of as the internal process of construction of meaning (in line with cognitive semantic approaches), is always a process of negotiation between the structure of our knowledge (linguistic and non-linguistic) and the organization of the language system, which is always maintained at the social/communicative level. Language is indeed cognitive, it is always a matter of cognitive processing internal to our mind, but at the same time it acquires a level of specificity that is established and maintained at the social level. Hence, the possibility of identifying a relational system of language and of linguistic meaning does not impose to think of the system as separated from the uses we make of language. On the contrary, the language system (*langue*) is a product of the standardization of the uses within a community of speakers. Let us recall that, according to De Mauro's reading of De Saussure, even in the Saussurian perspective, there is actually no strict distinction between the level of the system and the level of the concrete uses, but rather the language system arises (emerges) as a standardization of patterns of uses. *Langue* and *parole* are deeply intertwined: the language system emerges and is always dynamically restructured from language uses; in turn, language uses are grounded in and bounded by the structure of each system. Similarly, the representations we form of meanings in language are always cognitive (i.e. they are a product of the cognitive activity of individuals' multi-modal representations) and linguistics (i.e. they reflect

[132] We have extensively discussed this point in Chapter 2. Let us briefly recall that endorsing the principle of radical arbitrariness does not necessarily mean that language systems are not bounded to the way in which we perceive or categorize reality according to other modalities or that language is independent from our cognitive processes.

the bonds and conditions of a certain system of language). This allows us to possibly rethink the distinction between "evoked" and "encoded" meaning; if we assign the term "encoded meaning" to what is standardized at the level of the language system, and the term "evoked meaning" to the set of interconnected activated items in the mind of individual speakers, we could argue that linguistic units have "encoded meaning" in the sense that the "evoked meaning" is constrained by and always negotiated with the relational structure of the language system in the process of "doing meaning".

The claim of a usage-based language system of meanings does not necessarily incur into classical objections regarding the non-cognitive-plausibility of purely relational definitions of meaning inasmuch as, in line with distributional approaches (Lenci, 2008), the relational, "language-specific" level of linguistic meaning does not aim at giving an exhaustive and residual-free account of semantic (non-language specific) meaning nor at stating that such level is directly represented in the minds of speakers. In this sense, a usage-based language system is not a separate module of cognition, but rather an abstraction from the observable systematicity of language uses.

As we will see, this systematicity can be operationalized by means of the distributional semantic approach, which offers a precious perspective in our journey into the structure of language providing us with a set of methods to represent linguistic meaning.

For instance, distributional semantic models allow us to extract the relational system of language from large corpora, representative of language uses, and to represent it by means of the relations that linguistic items (represented as discrete or continuous vectors) entail in a semantic space (Boleda, 2020; Lenci, 2008, 2018). Furthermore, we will argue that, from a theoretical perspective, the distributional semantic hypothesis endorses a usage-based relational theory of linguistic meaning.[133] Distributional semantics assumes that the statistical distribution of words in

[133] Distributional semantics and constructionist grammar should not be considered as contrasting approaches, but rather as different levels of analysis of how language is extracted and constructed from language uses. Recent theoretical as well as empirical advances have shown that research perspectives at the intersection of distributional semantic and construction grammar offer rich lines of inquiry into the structure of our knowledge. Rambelli and colleagues (2019) offer a unified representation of grammar and meaning based on the assumption that language structure and properties emerge from language use and outline a first attempt to integrate semantic vectors derived from corpus data and formalized constructions. Indeed, the two frameworks share many theoretical assumptions and complement themselves on the basis of their respective strengths. However, we will not focus on this integration in our discussion, leaving it to a possibly subsequent research enterprise. On the contrary, we will here mostly draw from some of the differences the two proposal entail in the way they theorize meaning in the first place (a cognitive view versus a linguistic view of meaning), with the hope of showing that the two are indeed complementary and ought to be kept together if a unified notion of meaning wants to be reached.

linguistic context plays a key role in characterizing their semantic behavior. The theoretical foundation of distributional semantics is what has become known as the *Distributional Hypothesis* (DH): lexemes with similar distributional properties have similar meanings. Distributional approaches, a rich family of computational models, "raise the twofold question of the extent to which lexical properties can be reduced to their combinatorial behavior, as represented by their syntagmatic distribution in texts, and of the causal role of the contexts in which words are produced and observed in determining the structure and organization of semantic representations at the cognitive level" (Lenci, 2008, p. 2).

Indeed, the distributional hypothesis has important implications for the way in which we characterize semantic representations both at a linguistic as well as at a cognitive level. In fact, in its "strong" version (cf. Lenci, 2008) the distributional hypothesis is a cognitive hypothesis: the distributional behavior of a word in contexts and the repeated encounters with a word in different linguistic contexts are not only taken as a way to get at its semantic behavior in language use, but indeed as a way to explain its semantic content at the cognitive level. The encounters with words in different linguistic environments have an effect on the cognitive semantic representations of these words in, for example, our mental lexicon.

This leads the way to the possibility of re-discussing and re-framing the idea of a cognitive function of language. The structure of language, understood as the web of relationships observable in language usages starting from the distributional properties of words (or constructions) in linguistic context, becomes one of the ingredient by means of which we form our semantic (in the broad sense of the term as used in cognitive linguistic approaches) representations in our minds (Andrews et al., 2014; Dove, 2011, 2016; Louwerse & Jeuniaux, 2008, 2010; Vigliocco et al., 2009). Hence, the cognitive semantic system of meaning does not overlap and it is not limited to the linguistic structure of meaning, but rather the latter is thought of as an ingredient in the formation of the first. In conclusion, this proposal could allow us to rethink (at least at a theoretical level) a possible way in which *language might shape the way we think*.

In the next sections, we will proceed step by step and propose an in-depth discussion of each of the claims we are making about the possibility of identifying a "language-specific" level of meaning by means of the distributional properties of linguistic items, and of how such a "language-specific" level of meaning might in fact impact our cognitive semantic representations. In the first place, we will outline arguments to support our claim that distributional semantics offers

indeed a way of accessing a relational system of meaning, and that it is actually in line with the Saussurian tradition (§ 5.2). In the second place, we will briefly present distributional semantic models (DSMs) and discuss how they perform in both linguistic as well as cognitive tasks (the linguistic and the cognitive benchmark respectively) (§ 5.2. – for the cognitive benchmark see § 6.1.2). In the next chapter, we will turn to the analysis of distributional semantics in cognition; we will present the cognitive distributional hypothesis as developed within a psycholinguistic framework (§ 6.1) and we will review recent models of cognition that integrate embodied and distributional perspectives (§ 6.2). Lastly, we will present our own study concerning how the distributional properties of words are one important ingredient in the way in which we acquire new linguistic meaning (§ 6.3).

5.2. Uncovering the theory behind the methods

When we observe language is use, we notice a wide range of variability in the ways in which words combine, or in the ways in which we build sentences, we compose texts, or in which conversations unfold. Before even becoming a matter of cross-linguistic comparison, i.e. of how different languages combine words, structure sentences, and so on, variability is an intra-linguistic, even intra-individual phenomenon. Looking at how natural language unfolds in written and spoken utterances within a community of speakers, as well as at the individual level, may often give the feeling of being an unstructured series of idiosyncratic, individual uses, much like at the end of a circus performance, when all characters, athletes, animals, and so on, show up on stage giving bits and pieces of their actual complete range of abilities. Nonetheless, despite the considerable variation among individual instances and the apparently chaotic way in which units behave, there are considerable regularities and structured patterns in the way language unfolds over time and in the way in which linguistic units behave. In much the same way in which there is a choreography that underlies the movements of the very different kinds of circus characters on stage, individual uses of words and constructions exhibit both variation and structure.

From both a theoretical as well as a methodological standpoint, distributional semantics (DS) allows us to extract structure and organized behavior from the apparent chaotic mass of language uses. The primary focus of DS is lexical meaning: distributional semantic models (DSMs) offer an ever-growing set of methods to abstract and represent words meaning from language use, and more precisely from the linguistic context in which words occur. Moreover, DSMs learn representations of the meaning of words from language data, i.e. from text corpora – which can be taken as proxies of language use. Corpora are crucially connected to distributional semantics: as a repository of language uses, they represent the primary source of information to identify the word distributional properties. Nonetheless, the range of scopes and applications of DS is not limited to corpus linguistics, but rather allows to tackle issues and questions regarding the format and origin of words at a linguistic as well as at a psychological level.[134]

[134] Distributional representations are now mostly limited to the information provided by the linguistic contexts in which words occurs. However, this is not a restriction that arises from a limitation intrinsic to the distributional hypothesis, but it is rather due to the fact that the computational techniques to extract and model features form linguistic context are much more advanced that those to extract features from non-linguistic context. The narrow definition of context now in use among distributional semanticists does not prevent to admit a broader definition of context that might include other kinds of contextual information and a consequent development of models that are

5.2.1. Distributional semantics in a nutshell

The theoretical assumption underlying distributional approaches is that statistical distributions of words in context plays a key role in characterizing their semantic behavior. The distributional hypothesis (DH) states that lexemes with similar distributions have similar meanings: "[t]he degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear" (Lenci, 2008, p. 3). At least some aspects of the meaning of words (understanding which aspects of meaning can be accessed by means of distributional semantic representations is exactly the point of this chapter) depend on the distributional properties of words in the linguistic context in which they are observed. Consequently, the notion of *context* plays a major role in the theoretical foundation of DS and is one of the key parameters of distributional semantic models.[135] Together with the notion of context, the notion of *similarity* is a key theoretical component of distributional approaches to meaning: words are similar if they occur in similar linguistic context, i.e. there is a correlation between distributional similarity and meaning similarity, which allows us to utilize the former in order to estimate the latter. We will extensively discuss the notion of similarity later in this chapter given its crucial role in most linguistic and psychological investigations on meaning.

The distributional approach to natural language semantics takes advantage of the mathematical framework of vector spaces and linear algebra: vector spaces provide a natural mechanism for applying the geometrical concepts of distance and similarity to language data (S. Clark, 2015). The meaning of words can be represented and related geometrically: meaning is

able to embed information of different sources in the representation of meaning (Lenci, 2008). We will also see how this line of development (that is likely to lead to multi-modal distributional representations) is an active ongoing area of research in distributional approaches to language (§ 5.4.). Nonetheless, for the purposes of our discussion, the restriction of the notion of context to linguistic context is actually beneficial in the sense that DSMs trained on linguistic data allow us to tackle one specific aspect of linguistic meaning, i.e. the relational "language-specific" structure of meaning. This does not mean that distributional representations need to be limited to this; indeed, distributional representations are based on multiple assumptions, some of which (e.g. statistical learning) may indeed prove useful in modeling other aspects of human experience (e.g. visual and auditory experiences) and to provide representations that combine the different information provided within a broadly understood context.

[135] Here we take the notion of context as already limited to the linguistic context. Even within this limited definition, the notion of context remains problematic, non-uniform, and subject to considerable variations. For example, contexts can be whole texts, portions of texts, sentences, portion of sentences, and so on. The definition of context we are taking into consideration is not just a matter of theoretical debate, but it rather has direct methodological implications as it is one of the parameters on which DSMs are trained: different kinds of representations of words meaning and different scopes of the model depend on how we parametrize context in a given model (we will come back to this point in § 5.3.).

represented by means of vectors in a high-dimensional space, the fine-grained structure of which is provided by considering the context in which words occur in large text corpora.

Word-vectors can easily be compared for similarity in the vector semantic space, using any of the standard similarity or distance measures available from linear algebra (e.g. the cosine of the angle between two vectors).

Vectors are indeed the most useful numerical data structure that can be used to formalize contextual representations: the sequence of numbers forming a vector encodes the statistical association strength between a word and a certain context or distributional features. Since n-dimensional vectors represent the coordinates of points in a n-dimensional space, if we associate a word with a contextual representation and we formalize the latter as a vector, we can also conceive words as points in a "distributional space", i.e. a space whose dimensions are provided by the relevant linguistic contexts, and in which the position of a word-vector is determined by its statistical distribution in each context. As a consequence, similarities or differences between the words can be thought of as the relations entailed between word-vectors in a semantic distributional vector space.

Statistics is another key component of distributional semantics: it allows to abstract the relevant features of the context that enter into forming the contextual representation of words and to single out the relevant features to characterize a word distributional behavior. Lexical representations obtained by means of distributional semantics are context-based, distributed, quantitative and gradual: meaning derives from the way in which words interact in different linguistic contexts and are encoded in vector dimensions that represent the global distributional history of the word abstracted from the context of use, rather than some specific set of semantic features or meaning components (Lenci, 2008).

Is this enough to characterize what kind of semantics distributional semantics is? As a set of methods that share the assumptions and techniques described above, DS is primarily a quantitative method to represent and analyze linguistic meaning. However, as we have seen in the previous chapters, researches in both linguistics and psychology largely disagree on what counts as meaning or semantics, and on how to access and represent it.

Thus, it is more than fair to ask what kind of meaning DS is tackling on, what kind of distributional properties we should look for, and what kind of representation we obtain by means of DSMs. These questions are central for understanding the strengths and limitations of

distributional approaches and are at the core of the controversies about the place and role of distributional semantics in the study of meaning.

It has been argued against DS approaches - from both cognitive and formal linguistics perspective - that meaning cannot be explained in terms of language-internal word distributions, but needs to be anchored onto extra-linguistic entities, whether they be conceptual representations in the speakers' mind or objects in the world. In brief, distributional representations of word meanings essentially lack to address the referential aspect of linguistic meaning and they are intrinsically ungrounded (non-embodied) symbolic representations, since they only represent statistical distributional relations between symbols in the linguistic input, thus losing their cognitive plausibility within cognitive paradigms that view concepts, and hence meanings, as fundamentally grounded in human sensory-motor experience (we will come back to the discussion of these issues in § 6.2.2). These critiques would be appropriate if DS presented itself as a full-blown account of meaning (linguistic, conceptual, representational). Quite the opposite, we will argue that the claim of DS is far more specific and less accountable for such objections: in our proposal, DS describes and represents meaning at the level of the "language-specific" relational structure of meaning and this aspect of meaning is *one of* the ingredients in the formation of semantic (conceptual as well as linguistic) representations.

5.2.2. The historical roots of distributional semantics

In order to support such view of DS approaches, we need to unpack and reveal the possible implications of the distributional hypothesis, which are in turn connected to the twofold history of DS. The roots of the DH can be traced back to the work of the American structuralist linguistic Zellig Harris and to the birth and development of the "London School" of linguistics with John R. Firth, both of which, as we will see, entail very different but deep connections to the Saussurian and European structuralist background.

Let us begin by briefly summarizing what are the different points of view that the distributional hypothesis led to. Firth's work especially has been the source of a second cognitive-oriented offspring in DS that led to the formulation of a "strong" version of DH (Lenci, 2008), which is a cognitive hypothesis about the form and origin of semantic representations in the mind

of speakers (as we have sketched many times already, distributional information are thought of as one of the ingredients that contribute to the formation of our mental representations of meaning).

As for its linguistic formulation, the "weak" distributional hypothesis (Lenci, 2008) is essentially a method for semantic analysis: the weakest assumption is that lexical meaning (however defined) determines the distribution of words in context and that its semantic properties, investigated by inspecting a relevant number of distributional contexts, act as constraints governing their syntagmatic behavior. In this sense, there is only a correlation between semantic content and linguistic distribution and meaning is a sort of "hidden variable" (for a review cf. Acquaviva, Lenci, Paradis, & Raffaelli, forth) responsible for the observable distribution.

A third possibility explored in the DS literature (that, as a matter of facts, can be understood as placing itself in between the very weak and the strong version) is that distributional methods based on the distributional hypothesis allow us to extract and identify the paradigmatic and syntagmatic relationships within the language system (Sahlgren, 2006, 2008), i.e. the *value* of the linguistic units. According to this last perspective, DS allows us to represent the relational and differential value of the units in the system as they become standardized in the language in use, and to tackle onto the "language-specific" level of meaning.

According to the theoretical perspective outlined throughout our discussion, asking whether distribution shapes meaning or if words have different distributions because they have different meanings (Palmer, 1976) is a reductive, still dichotomic, way to pose the question. If the perspective we are endorsing is that of a usage-based "language-specific" level of meaning, we would then argue that there is no supremacy or priority of one aspect (distributions) over other aspects of meaning, but rather that the two are inherently related and always intertwined in the way in which we do meaning. The way in which we do meaning is always a product of the distributional constraints, which we could think of as the systematic aspect of meaning, addressed by distributional semantics, and other kinds of cognitive, idiosyncratic, social aspects that constantly bend and reshape the structural "language-specific" level of meaning. At least at the current stage of research, DS is primarily a way to tackle onto the relational aspect of meaning, its procedures and assumptions stemming from the structuralist tradition.

The distributional hypothesis is rooted in the structuralist (and, we will argue, especially Saussurian) soil: on the one side, the works of Harris and Firth, the two founders of DS, bear

important relations (of both alignment and critique) with the structuralist tradition and, on the other side, the kind of meaning extracted by means of DSMs is inherently differential and relational.

The distributional methodology is often motivated by the work of Harris, who argued that "difference of meaning correlates with difference in distribution" (Z. S. Harris, 1954, p. 156). Harris's view was deeply embedded in the North-American structuralist paradigm exemplified by the work of Bloomfield, who strongly advocated for the need to eradicate the study of meaning (semantics) from linguistics to ensure the focus on the formal regularities of language addressed by rigorous methodological practices. In a similar way, Harris was primarily interested in establishing a linguistic methodology that could deal with what is internal to language. The distributional methodology provides exactly the formal framework according to which facts about meaning can be measured and quantified: "If we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C" (Z. S. Harris, 1954, p. 156). Hence, meaning becomes "part of the linguistic science, at least for those aspects that can be defined through the very same method with which any linguistic entity should be defined, namely distributional analysis procedures" (Lenci, 2008, p. 4).

The procedure advocated by Harris for the study of meaning in language is based on the structuralist intuition that meaning is inherently differential and that the linguistic analysis of language should be limited to what is internal to language, e.g. the combinatorial restrictions of linguistic entities. The primary focus of the distributional analysis is on the relationships entailed between words at a syntagmatic level: the relationships that words have with other elements in the stretch of language in which they occur. Distributions, therefore, would deal with purely linguistic relations of a syntagmatic kind (Palmer, 1976). The consistent core idea in Harris's thought is that linguistic meaning is differential (in the sense described above more than in a strictly Saussurian sense) and not referential, that differences of meaning are mediated by differences in distributions (Sahlgren, 2008).

According to Sahlgren (2006, 2008; but see also Lapesa, Evert, & Schulte im Walde, 2014), rethinking the distributional approach in connection to, on the one side, the Harrisian methodological formulation and, on the other side, the Saussurian relational view of meaning (in terms of paradigmatic and syntagmatic relations) grants us the possibility to reformulate and precisely state the kind of meaning that is represented - and allowed - in distributional

representations: "a distributional model accumulated from co-occurrence information contains syntagmatic relations between words, while a distributional model accumulated from information about shared neighbors contains paradigmatic relations between words" (Sahlgren, 2008, p. 7).[136]

However, pointing back to Harris's formulation of distribution, and hence to the American structuralist tradition, does not allow us to grasp the full reach of the DS approach in the study of language. Distributional semantics allows us to deal with the vagueness, incompleteness, and fuzziness of linguistic meaning in a theoretically strong way as well as with the more strictly structural aspects of meaning (being the two inherently intertwined and interdependent). There is a long and withstanding tradition in linguistics that views the incomplete, noisy and imprecise form of natural language as an obstacle that obscures rather than elucidates meaning. It is very common in this tradition to claim that we therefore need a more exact form of representation that obliterates the ambiguity and incompleteness of natural language. On the contrary, distributional semantics, in line with proponents of descriptive approaches in linguistics, argues that ambiguity, vagueness, polysemy, incompleteness are the essential properties of natural language and can be addressed by a relational view of meaning grounded in the Saussurian perspective on language. DS combines the needs of structuralist approaches focused on the linguistic relationships between the linguistic units within the language system with the perspectives of usage-based approaches, according to which the use we make of the linguistic units (lexemes or constructions) is a key aspect for understanding how we construct meaning at a linguistic as well as at a cognitive level. As we have mentioned before, and we will see in more details later, the strong version of DS may have important implications for the way in which we understand meaning not just at a linguistic level, but at a cognitive level as well.

5.2.3. Firth and the structural perspective on meaning

The need to expand the study of meaning to levels beyond the systemic relations was already evident in the second, very often cited, founder of the distributional hypothesis, the British linguist John R. Firth. If Harris provided the overall methodology for the semantic analysis

[136] We will become more familiar with the notion of *neighbor* in the next section. As for now, neighbors are words that tend to occur in the same linguistic context. The extraction of neighbors, and hence of paradigmatic associations, is now the main operation carried on by DSMs.

developed by DS, Firth offered a wider theoretical framework for the study of meaning and traced multiple lines of inquiry that have become fundamental in DS analyses. Firth's most cited idea is that "You shall know a word by the company it keeps" (Firth, 1968a, p. 179, first published 1957). For Firth, this "keeping company" was just one part of the meaning of words; meaning was to be found in the context of situation and in many other levels of analysis as well. He called them "modes" of meaning (Firth, 1957b, first published 1951): Firth's contextual theory of meaning was based on the assumption that meaning is a very complex and multifaceted reality, inherently related to language use in context (e.g. social settings, discourse, etc.), and that it had to be studied at multiple levels of analysis. One of the key "modes" of meaning of a word is what he called "meaning by collocation", determined by the context of surrounding words. Collocations are abstractions at the syntagmatic level; they are not mere juxtapositions of words, but actual ordered *mutual expectancies*. According to Firth, collocations are one of the constitutive aspects of meaning.

The notion of meaning and the scope of descriptive linguistics, a branch of general linguistics, proposed by Firth deserve a more specific discussion. According to Firth, descriptive linguistics is one aspect of general linguistics, i.e. it is one of the applications of a theory of language. The main concern of descriptive linguistics is "to make statements of *meaning* in purely linguistic terms, that is to say, such statements are made in terms of *structures* and *systems* at a number of *levels of analysis*: for example, in phonology, grammar, stylistics, situation, attested and established texts" (Firth, 1968b, p. 97)[137]. Meaning is a complex and multifaceted reality that is always the product of the mind and the body, the nature as well as the nurture of human beings, their social and cultural environments. The role of descriptive linguistics is to isolate meaning in linguistic terms, although completely rejecting the ontologization of the descriptive analysis. "To make statements of meaning in terms of linguistics, we may accept the language event as a whole

[137] Firth descriptive linguistics is essentially a corpus-based linguistics grounded in the notion of context of situation. The notion of *context of situation*, drawn from Malinowski (which regarded the context of situation as a sort of behavioral matrix in which language had meaning), is used by Firth as a schematic construal to apply to language events which brings into relation categories such as the relevant features of participants, their verbal actions as well as their non-verbal actions, the relevant objects, and the effects of the verbal action (Firth, 1957c). Any text can be regarded as a constituent of a context of situation or of a series of such contexts; "the important thing to remember in this approach is the abstract nature of the context of situation [...]. Instances of such context are attested by experience. The context of situation according to this theory is not merely a setting, background or 'backdrop' for the 'words'. The text is the focus of attention on renewal connection with an instance in regarded as an integral part of the context, and is observed in relation to the other parts regarded as relevant in the statement of the context" (Firth, 1968a, pp. 175–176).

and then deal with it at various levels, sometimes in a descending order, beginning with social context and processing through syntax and vocabulary to phonology and even phonetics, and at other times in the opposite order" (Firth, 1957b, p. 192).

Hence, meaning is everywhere in language in the sense that it is dispersed at a series of congruent levels of analysis, which taken together form a sort of linguistic "spectrum" (Firth, 1957a); in this spectrum, the meaning of the whole linguistic event is dealt with by a hierarchy of linguistic techniques descending from social contextualization to phonology. Hence, meaning is essentially a property of the mutually relevant people, things, and events in the situation. Interestingly, in Firth's proposal, meaning should not be regarded as conceptual: the meanings of words are not immanent essences or detachable ideas which we could traffic in. Linguistic descriptions should avoid the temptation of reifying individual words as isolates of conceptual meaning or as "naked ideas"[138] (Firth, 1968c): "words must not be treated as if they had isolate meaning and occurred and would be used in free distribution. A multiplicity of *systems* derived from carefully contextualized *structures* would seem to be indicated. The *structures* attributed to 'texts' are not to be given ontological status. They are schematics" (Firth, 1968d, p. 18). Similarly, meaning should not be thought of as a relation in the mind between facts and events (referents) on the one hand and symbol or words we use to refer to them on the other, but rather as situational relations in a context of situation. Firth's contextual theory of meaning emphasizes the interrelations of the terms in the actual observable context itself.

Starting from this general orientation, descriptive linguistics - operating within a contextual theory of meaning - employs abstractions that allow us to handle language in the interrelated processes of personal and social life in the flux of experience: "we study the flux of experience and suppress most of the environmental co-ordination of what we examine, regarding the essentials as *instances* of the general categories of the schematic constructs we set up. We see structure and

[138] In discussing the problem of translation and the use of the statement of meaning by translation as part of the linguistic analysis, Firth pointed out the problem of "building bridges" between languages by means of "naked ideas", a sort of stable conceptual refence to evaluate whether two words are the same or not; "all linguists must, of course, make use of common human experience and even all the highly abstract concepts of the common human situation, in order to bridge the gulf separating one language from the other. But they must always bear in mind that the English word 'kindness', for example, does not represent a 'naked idea' of any value to linguistics, and in that sense no other language in the world has a word for 'kindness'. It is equally unprofitable to make statements such as, this language has no word for 'the' or 'lamb', or that it has no verb 'to be'" (Firth, 1968c, p. 76). In turn, Firth proposed a polysystemic linguistic analysis, i.e. a comprehensive description of a given language, as a profitable way of reaching translatability.

system as well as uniqueness in the instance, and its essential relationship to instances other than itself" (Firth, 1968d, p. 16). Meaning is inherently relational in the sense that, "in dealing with language in the matrix of experience", meaning has been "split up and regarded as a relation or system of relations" (Firth, 1968a, pp. 173–174).[139] We can identify two main set of relations: firstly, the relations interior to the text and, secondly, the set of situational relations, which are set up between the text and the other constituents of the situation. The interior relations are of two kinds, syntagmatic and paradigmatic, from which the distinction between structure and system follows. Structure, whether it be phonological or grammatical, is a syntagmatic relationship. It is an interrelationship of elements within the text, whereas a system such as a system of vowels or of grammatical forms is in the nature of a paradigm. The values of a paradigm are determined by the interior relations within the paradigm itself. Systems and structures are essentially an abstraction from language use. Hence, "linguistic analysis states the interrelation of elements of structure and sets up system of 'terms' or 'units' and end-points of mutually determined interior relations" (Firth, 1968a, p. 177).

Within this framework, Firth seems to draw the notion of *value* from De Saussure: meaning has to be described as a relational network of structures and systems at clearly distinguished but congruent levels, converging in renewal of connections in experience. Meanings, as described at a linguistic level, are not direct mappings to the external or conceptual organization of our experience, but rather it is in the relationships between the units of the system that we see meaning arising. In this sense, for example, *notional time* and *grammatical time* may not have equal systems; "[n]otional time, generally speaking, is a different thing altogether from grammatical time, which differs from language to language. Grammatical time is not limited to or bounded by the speaker's temporal world but has unlimited range, always, however, within the verbal time-resources of the given language. Each language has its own means of handling 'experiential' time,

[139] It seems important to point out that the way in which Firth used the term "relational" does not completely overlap with the Saussurian or structuralist notion of "relational meaning". We have seen how, according to Firth, meaning is relational in a much broader sense that in the structuralist perspective: meaning is the product of relations between various and multiple aspects of our experience, each of which can be described by means of the abstractions of systems and structures. Hence, the linguistic description of language is inherently polysystemic in the sense that it is a combination of systems at every level of analysis and between different levels. Nonetheless, as we will see when analyzing the system of collocations, each system is in itself relational, in that the elements that compose the system acquire their value by means of the relations internal to the system. Within each level, meaning seems to be regarded as relational in the Saussurian sense, but, at the general level of description, language is not a system "*où tout se tient",* but rather inherently polysystemic.

has its own 'time-camera' so to speak, with its own special view-finders, perspectives, filters and lenses" (Firth, 1968a, p. 188). Similarly, "[a] singular in a two number system has different grammatical meaning from a singular in a three number system or a four number system such as in Fijian which formally distinguished singular, dual, 'little' plural and 'big' plural" (Firth, 1957a, p. 227). Even more to the point of our discussion, the various formal structures and collocations of a given word can be regarded as criteria for setting up a system of distributed variants (senses), in which each variant is a function of the others and of the whole system. For example, in regard to the analysis of the word *get,* Firth states that "[e]ach sectional definition in shifted terms is *not* a statement of a concept or of the 'essential' meaning of the word *get* itself, but a descriptive indication of the relation of the collocations to generalized contexts of situations" (Firth, 1968d, p. 20). Hence, systems of collocations, i.e. systems of distributional patterns of uses and mutual expectancies, can be essentially regarded as systems of sense.

In conclusion, according to Firth, language is systematic at every level of analysis and when studying any given language, we need to identify the specific language (polysystemic) system, actively maintained by people carrying and conveying the system or systems. Each speech event in a context of situation, as a technical abstraction from utterances and occurrences, is always an expression of the language system from which they arise and to which they are referred (Firth, 1957d).

The last central aspect that we want to point out about Firth's proposal is that it can essentially be regarded as a usage-based systemic view of meaning with a specific interest in describing the system of language from a linguistic perspective, within which collocations are one of the linguistic "modes" of meaning. Firth emphasizes the role of the individual as a member of the culture and of the society in the study of meaning and assigns a pivotal role to *parole*, the individual acts of speech that actively and consistently maintain the language system. "Language [...] is everywhere actively maintained by *persons*, that is, by people who are members of society. Language in this sense must be *systemic*, because it owes its genesis and its continuance to human bodily systems living in society" (Firth, 1957c, p. 187).[140] In this sense, the systemic aspects of

[140] This idea rests on a fairly complex notion of personality intertwined with the notion of language. Speech, both oral and written, is an outcome of personality and language, which are in turn both seen as a *binder of time*, of the past and the future in the present. "On the one hand there is habit, custom, tradition, on the other innovation and creation. Every time you speak you create anew, and what you create is a function of your language and your personality. From that activity you may make abstraction of the constituents of the context, and consider them in their mutual relations" (Firth, 1957d, p. 142). Both language and personality are fundamentally grounded at the level

language, the Saussurian *langue*, cannot be separated from the acts of speech of the individuals, which are grounded in the physical and social milieu[141]; language is a vector "of the continuity of repetitions in the social progress, and the persistence of personal forces" (Firth, 1957c, p. 183).

Firth's main concern was to deal with speech events, maintaining that these events are expressions of the language-system(s) from which they arise and to which they are referred as much as of the individual "personality" of the language users. Texts, as contexts of situations, are the fundamental place to look for patterns of uses, hence for the systematicity that arises from and subtends those uses; although speech events are unique, they nonetheless have features in common with other speech events.

Collocations, together with colligations, seem to be one of the main patterns in which we observe systematicity and a fundamentally relational view of meaning: "[i]t can be safely stated that part of the 'meaning' of *cows* can be indicated by such collocations as *They are milking the*

---

of the physical body, which already exhibits that systematicity that propagates at the level of personality and language.

[141] Firth's critique to the Saussurian notion of *langue* is openly stated in many passages in Firth's writings. Firth considered himself "by no means a Saussurian", even though, as we have seen, there are many evident points of contact between the Saussurian and Firthian positions, such as the need for a linguistic perspective in the study of meaning, the notion of value, and the focus on the relational aspects of the language system. Firth is particularly severe in his critique to the post-Saussurian structuralism, which he sees as a very limited approach and of which he openly rejects the reification of the system. According to Firth, the systems and structures of language are not a separate ontological entity, but rather a linguistic fiction, a descriptive scaffold that employs all technical recourses systematically for multiple statements of meaning in the appropriate linguistic terms (this is what Firth calls "structural linguistics" as opposed to "structuralism") (Firth, 1968e). As for Firth's relation with the Saussurian perspective, it is not difficult to spot three souls in Firth's work: the openly anti-Saussurian, the non-Saussurian, and the actually Saussurian (Koerner, 2000). Firth's critique to the dichotomic view of *langue/parole* is indeed the theoretical aspect in which the contrast is the sharpest. In Firth's view, De Saussure envisages a *langue* without *parole* (Firth describes De Saussure's position with the equation: *langage* minus *parole* equal *langue*). Even though he recognized the fundamental role of the *masse parlante* for the language system, Firth criticized the idea that this *masse parlante*, as a "collective consciousness" in which language as a function of the speaking mass in stored, is still regarded as on a different plane from individual phenomena. Hence, the main critique to Saussure is directed towards the isolation of the system of the *langue* from the individual bodily, psychological, and social life. In contrast to Saussure's position, Firth declared his fundamental interest for the study of speech, of language in use, and hence of the *persons* always and necessarily involved in any concrete act of speech: "[w]e are now a long way from De Saussure's mechanistic structuralism based on a given language as a function of the speaking mass, stored in the collective consciousness, and from the underdog, considered merely as the speaking subject, whose speech was not the 'integral and concrete object of linguistics'. The unique object of Saussurian linguistics is '*la langue*', which exists only in the *collectivité.* Now it is at this point that I wish to stress the importance of the study of persons, even one at a time" (Firth, 1957c, p. 183). As we have already argued in the previous chapters, this strictly dichotomic view of Saussure's *langue/parole* relationship is more a heritage of the structuralist interpretations of Saussure's thought than the actual Saussurian position. The most recent rediscovery of Saussure's idea of *langue* as always grounded in the acts of *parole* would allow us to rethink the theoretical relationship between Firth and Saussure as well (for a review on the relationship between De Saussure and Firth cf. Koerner, 2000; Love, 1986).

cows, *Cows give milk*. The words *tigress* or *lioness* are not so collocated and are already clearly separated in meaning at the *collocational level"* (Firth, 1968a, p. 180). The analysis of collocations reveals important aspects of linguistic meaning: the meaning of a linguistic unit is, at least to some extent, differential and it is the outcome of a local balance of relationships internal to the system. The value of a linguistic unit is determined by its position in the syntagmatic axis, i.e. by the words that surround it in the linguistic context. Furthermore, the analysis of the syntagmatic relations allows us to extract relationships at the paradigmatic level: in distributional terms, words that co-occur can be seen as entailing a syntagmatic relation, whereas words that share neighbors can be regarded as forming a paradigm (Sahlgren, 2006, 2008). Hence, language can be described, in a Firthian sense, as a polysystemic system of relations grounded in the language uses, which show both systematicity and uniqueness.

This brief digression into the historical roots of the distributional hypothesis allows us to draw some conclusions and to shed light, by means of the historical foundation of DS in the Saussurian perspective, on the kind of meaning address by distributional approaches to language. DS is grounded in the relational view of meaning in the sense that at least a part of the meaning of words can be regarded as the outcome of the local balance between elements internal to the system. The analysis of words in their context of use allows us to extract the relational aspect of meaning, e.g. to identify (and, as we will see, to quantify) the position of a linguistic unit in the paradigmatic and syntagmatic axes. Meaning by collocations does not reinstate a categorical view of meaning, but rather represents meaning at the linguistic level as a conglomerate of senses sensitive to contextual usage-based shifts. On the one side, distributional representations of linguistic meaning operationalize and mathematically encode the assumption that meaning (and meaning similarity) is a function of the distributional properties of words in context as well as a non-categorical view of meaning: distributional vectors are distributed representations statistically recording co-occurrences of lexemes with linguistic contexts (Lenci, 2018). Symbolic semantic representations are discrete and categorical. Distributional representations are instead graded and distributed, because information is encoded in the continuous values of vector dimensions. On the other side, distributional models allow us to effectively grasp the ambiguity and multiplicity of the senses as they are instantiated in language use (Schütze, 1997; Sinclair, 1991). Furthermore, distributional semantics considers the meaning of a lexical item as dependent on its relations with the other lexemes in the semantic space. Semantic spaces (as an expansion and operationalization of the

notion of lexical field typical of structural linguistic), extracted by means of distributional analysis as the set of words that are in paradigmatic and syntagmatic relations, are not bounded to specific delimitations and boundaries posed by the researchers; semantic spaces arise from the interplay between syntagmatic and paradigmatic dimensions of meaning and can hence be studied as grounded in linguistic evidence (Apresjan, 1966).

DS is inherently a usage-based approach to meaning, according to which the source of the systematicity in language is not an a priori established language system, but rather the standardization of patterns of uses that are at the same time the product of the individual, idiosyncratic, encyclopedic uses and of the boundaries historically and socially established which gave rise to and constantly maintain the language system. As a consequence, the language system is always dynamic in the sense that it is always subjected to the stretches made in the actual uses of language; the language system is the outcome of the balance between constant stability and constant change.[142]

These very general observations give us a first intuition of what DSMs can tell us about how language is structured starting from the examination of language in use. Our aim in this section was to show the deep connection between the Saussurian perspective and the distributional semantic approach in its historical roots as well as to offer a first outlook on how some of the fundamental themes of the Saussurian tradition find room for debate and evidences in contemporary distributional approaches. In regard to the former, we have seen how both the Firthian and the Saussurian tradition (although with substantial differences) highlight the need to view meaning as (at least in part) a linguistic fact: in our terms, the "language-specific" level of linguistic meaning can be identified in the relations that the linguistic units entail within the language system, which is always the outcome of the stability provided by the system itself and the variability of the individual uses in the acts of speech. Distributional semantics is in line with these theoretical assumptions and allows us to find structure in uses providing the methods by means of which we can describe the relationships internal to the language system. In regard to the latter, the next section will provide more details about what the different types of DSMs allow for in terms of, for example, individuating different senses of a word, differentiating between semantic relations, representing constructions, and so on.

[142] For example, distributional semantics allows us to tackle onto language change and onto how standardized meaning evolved under the pressure of uses (e.g. Hamilton, Leskovec, & Jurafsky, 2016; Rodda, Senaldi, & Lenci, 2016).

5.3. The theory in practice: Distributional Semantic Models (DMSs)

The aim of this section is to provide the reader with a general outline of the types of models available in distributional approaches, of what kind of systemic relations can be identified by means of the various models, and to present how they perform on some main linguistic benchmarks.[143] The literature on distributional models is vast and increasingly growing, given the interest that distributional semantics is receiving not just in linguistic and psychological disciplines, but also because of its efficiency in developing new methods in Natural Language Processing, especially related to deep learning and machine translation. The availability of big corpora of language data in a variety of languages and the advancements in computational methods to extract information from these corpora and to analyze them are at the center of recent developments in the field and constantly lead to new questions that can be asked and solved by means of new computational implementations of distributional models.

Given the vastness of the research questions and implementations that make use of distributional approaches to language meaning, our aim is not to cover them all or to offer any new insight and perspective. We aim at presenting a general outline of the possibilities offered by distributional models in understanding the structure of linguistic meaning and of their application in the study of language in cognition. Even within the boundaries of this restriction of our interest to the linguistic and psychological literature, the range of topics that could be covered in a review in extremely vast. The way in which distributional models are built offers a first ground for important variations in both the modelling aspect as well as in the range of linguistic and psychological issues that can be accounted for.

A distributional semantic model is a particular configuration of the parameters used to build distributional representations. The parameters to be determined when building distributional representations include the selection of target lexemes, the definition of the context type, the choice of weighting scheme, the application of dimensionality reduction, and the choice of a vector similarity metric. Hence, it seems clear that there can be as much variation in the models as the possible combinations of the different parameters. A lot of work has already been done in experimenting with parametrical variations and in analyzing the outcomes of the models (for

[143] Another set of benchmarks for DSMs is of a cognitive kind: DSMs outputs are most often evaluated in how they perform in predicting linguistic behavioral data. We will focus on these and other related issues in § 6.1.2.

important and comprehensive reviwes see S. Clark, 2015; Erk, 2012; Lenci, 2018; Turney & Pantel, 2010).

As to how to evaluate the output of DSMs, it is important to point out that the implementation of distributional models is strictly connected to the research question at task and distributional models can be evaluated on the basis of their performances in a variety of linguistic and psychological tasks (e.g. Kiela & Clark, 2014; Lapesa & Evert, 2014). In this sense, there is no such thing as "the best model" overall, but only models that perform better or worse in a given task and under certain parametric conditions. However, this should not be seen as a limitation, but rather as a strength of the field; distributional semantics is an extremely active field of research that is constantly reshaped by the question asked and the methods that become progressively available. Beyond this variability, the implementation of distributional models rests on the theoretical assumptions we outlined in the previous sections and allows us to address both relational and dynamic aspects of linguistic meaning.

In this section, we will briefly discuss some of the main differences in DSMs based on the presentation and discussion of their constitutive parameter and underlying methods. We will discuss the notion of similarity and relatedness in the DS framework, focusing on what kind of semantic relations can be extracted by means of distributional representations, and we will briefly review some of the new frontiers of distributional modeling in accounting for the structure of language.

5.3.1. What DSMs are: the two general pipelines for constructing DSMs

How do distributional models work? At a very general level, we could say that distributional models are a set of computational procedures to extract vectors of meaning, hence distributional representations, from (generally large) linguistic corpora. The structure of the model comprises a set of parameters, among which the most relevant is the definition of the context: the type of context (text versus words), the size of the context (whole texts or context window of various sizes), whether we take into consideration syntactic dependences or not. The kind of distributional representations (the output of the model) we obtain largely depends on how we parametrize the notion of context in a given model.

A second fundamental difference among the models, which again gives rise to differences in the representations we obtain, is the method used to build distributional vectors: models can either count the frequency with which words co-occur (both if we consider the context as a text region or as a word-to-word co-occurrence) or they can learn to predict words from the words surrounding them (count versus predict models) (Baroni, Dinu, & Kruszewski, 2014; Lenci, 2018).

The method by which models are trained allows us to distinguish between implicit and explicit distributional representations (Levy & Goldberg, 2014c, 2014b). Distributional representations obtained with count models are vectors that encode the distribution of word co-occurrences in a given context: the frequency of co-occurrences between target words and context (words or text) is counted and represented in a co-occurrence matrix (word-by-word matrix or word-by-region matrix). Raw frequency counts tend not to be very informative given that the distribution of frequency of lexemes follows Zipf's law (cf. Piantadosi, 2014) and that, consequently, these vectors are highly dimensional, as the number of contexts in language data tend to be very large.[144] Hence, vectors solely based on frequency counts tend to be very sparse (most of their dimensions are zero) and can be referred to as explicit vectors (Lenci, 2018).

Because of the skewed data distribution, vectors based on raw counts tend to miss some important generalizations in distributional data, and especially the abstract semantic structures that is latent in the raw data collected for word co-occurrences. In order to reduce the noisiness of the data and to extract the latent structure, matrix dimensionality-reduction techniques are used to map the data in the high-dimensional space of linguistic contexts to a space of fewer latent dimensions. This process is also called feature extraction, because the dimensions of the reduced space are new features extracted from the original data. Thus, instead of representing target lexemes using the linguistic contexts they co-occur with, we represent them in a latent semantic space of implicit vectors with a much smaller set of abstract features discovered in distributional data.[145] By means

[144] Distributional representations use various forms of weighting functions to overcome the problems of raw frequencies and to assign higher weights to co-occurrences that are more informative about the content of the target lexemes. The most common weighting function in DS is positive pointwise mutual information (PPMI). PPMI measures how much the probability of a target–context pair estimated in the training corpus is higher than the probability we should expect if the target and the context occurred independently of one another (Bullinaria & Levy, 2007).

[145] Even though they are generally referred to as *latent features*, the latent dimensions extracted by means of dimensionality-reduction techniques do not represent clear and specific semantic values, but rather have to be thought of as dimensions in a vector space. Hence, vectors extracted with reduction techniques are not directly interpretable, in the sense that there is no one-to-one correspondence between features and linguistic contexts.

of dimensionality-reduction techniques we obtain dense vectors, generally addressed as implicit vectors, in which most of their dimensions are non-zero (Lenci, 2018). The most common technique to extract implicit vectors from count data is Singular Value Decomposition (SVD), which is a statistical technique used to map the co-occurrence matrix onto a reduced latent semantic space with a matrix reduction algorithm.[146] Count models that make use of dimensionality reduction techniques are, for example, Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), which starts by counting how many times a word is observed within a document or a paragraph, and the Hyper-space Analogue to Language (HAL; Lund & Burgess, 1996), which counts how many times words co-occurred in a relatively narrow sliding window, usually consisting of up to ten surrounding words.

A second, more recent, family of models is based on the notion of prediction: these models implicitly learn how to predict one event (a word in a text corpus) from associated events. Instead of first collecting context vectors and then reweighting these vectors based on a certain criterion, the vector weights are directly set to optimally predict the contexts in which the corresponding words tend to appear.[147] Since similar words occur in similar contexts, the system naturally learns to assign similar vectors to similar words. Instead of counting co-occurrences, prediction DSMs are neural network algorithms that directly create low-dimensional implicit distributional

Similarly, as we will see, neural word-embeddings are considered opaque, in the sense that it is hard to assign meanings to the dimensions of the induced representation (Levy & Goldberg, 2014a).

[146] For other types of weighting functions and matrix dimensionality-reduction techniques see Curran (2003), Evert (2008), Turney and Pantel (2010), Kiela and Clark (2014).

[147] The semantic content of a word therefore lies in its global distributional history encoded in the vector, rather than in some specific set of semantic features or relations. Neural networks are general algorithms that encode information with vectors of neural unit activations and learn high-order representations from co-occurrence statistics across stimulus events in the environment. Connectionism is fully consistent with the distributional hypothesis, since linguistic co-occurrences are just a particular type of stimuli that can be learnt by neural networks. Furthermore, it has been argued that "[e]ven though the predict models originated outside the context of psychological research and were not concerned with psychological plausibility, the simple underlying principle – implicitly learning how to predict one event (a word in a text corpus) from associated events – is arguably much better grounded psychologically than constructing a count matrix and applying arbitrary transformations to it" (Mandera et al., 2017, p. 58). Indeed, predict models (similarly to what we have seems for connectionist models in general) are able to implicitly learn word meanings from raw linguistic data and to store knowledge about language in their hidden layers. Nonetheless, at least at the current stage of research, it seems premature to certainly state the cognitive plausibility of computational models, especially when addressing how humans learn linguistic meaning from language data. Furthermore, the cognitive plausibility of predict models (as well as of other models such as Long short-term memory (LSTM) models based on advancements in deep learning techniques) is more of interest for research in natural language processing (NLP) than in distributional semantics, in which no argument is made about the cognitive mechanisms that underlie the acquisition of meaning from language data.

representations by learning to optimally predict the contexts of a target word. These representations are also commonly referred to as (neural) word embeddings.[148] Instead of explicitly representing the words and their context in a matrix, the model is based on a relatively narrow window (similar in size to the one often used in the HAL model) sliding through the corpus. By changing the weights of the network each time an input is encountered and processed, the model learns to predict the current word given the context words (Continuous Bag of Words model; CBOW) or the context words given the current word (skip-gram model)[149] (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013; for a review: Y. Goldberg & Levy, 2014). These models are implemented in the word2vec software.[150]

A key difference between count models such as LSA and word2vec is that the latter produces latent representations with the useful property of preserving analogies, therefore indicating appealing linear substructures of the word vector space, whereas the former takes better advantage of the overall statistical information present in the input documents. However, the advantage of either approach is the drawback of the other. GloVe (Global Vectors) addresses this issue by performing unsupervised learning of latent word vector representations starting from global word–word co-occurrence information (Pennington, Socher, & Manning, 2014). While both word2vec and GloVe are popular approaches, a key difference between the two is that the former

[148] The quantitative and gradable character of distributional representations makes them very similar to the way information is represented in artificial neural networks. Connectionist models use non-symbolic distributed representations formed by real-valued vectors such that "each entity is represented by a pattern of activity distributed over many computing elements". Distributional representations are also distributed because the semantic properties of lexical items emerge from comparisons between their n-dimensional vectors, for example by measuring their similarity in distributional vector space. "The notions of distributed and distributional representations are closely related but need to be kept well distinguished. In fact, the former concerns the way semantic information is represented with vectors, while the latter concerns the source of the information used to build the vectors. The term 'distributional' specifically refers to the property of vectors to encode the statistical distribution of lexemes in linguistic contexts. *All distributional representations are distributed, but not all distributed representations are distributional*" (Acquaviva, Lenci, Paradis, & Raffaelli, forth).

[149] The architecture of predict models in very similar to the one we already discussed for Simple Recurrent Networks. Both the CBOW and the skip-gram models are simple neural networks consisting of an input, a hidden and an output layer. In the input and the output layers each node corresponds to a word. So, the number of nodes in these layers is equal to the total number of entries in the lexicon of the model. The number of nodes in the hidden layer is a parameter of the model. The training is performed by sliding a window through a corpus and adjusting the weights to better fit the training examples. When the model encounters a window including a phrase black furry cat, the CBOW model represents the middle word furry by an activation of the corresponding node in the output layer and all context words (black and cat) are simultaneously activated in the input layer. Next, the weights are adjusted based on the prediction error. In the case of the skip-gram model the association between each of the context words (black and cat) is predicted by the target word (furry) in a separate learning step. When training is finished, the weights between the nodes of the input layer and the hidden nodes are exported as the resulting word vectors.

[150] code.google.com/p/word2vec/

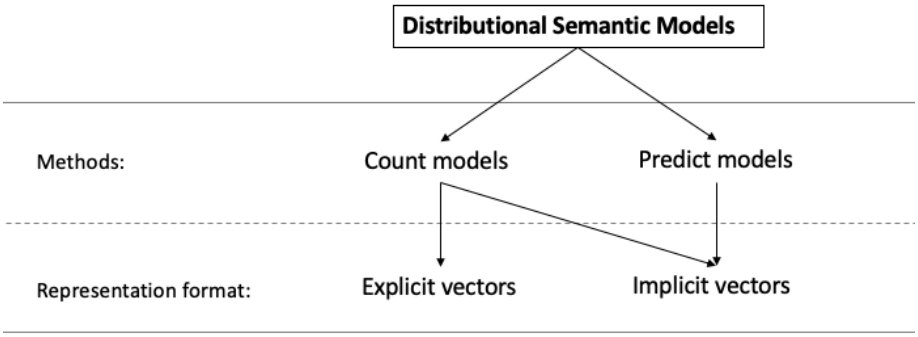


Figure 19. Types of models based on the method to construct them (adapted from Lenci, 2018)

is a predictive model, whereas the latter is a count-based model (see figure 19 for a recap on the different families of DSMs).

For both count and predict models, the tuning of the internal parameters affects both the kind of linguistic regularities that can be extracted as well as the performance of the models in predicting linguistic and behavioral data. In principle, there are no theoretical reasons to prefer one kind of settings compared to another; hence, there are many empirical studies that programmatically address the evaluation of parametrical variations and compare the outcomes of models on various tasks (e.g. Kiela & Clark, 2014; Lapesa & Evert, 2014). Furthermore, when comparing count models and predict models in general, little agreement is reached on whether a family of models actually outperforms the other. Even if predict models have shown to do better in a multiplicity of tasks (Baroni et al., 2014), both kind of models, when properly tuned, are able to capture the same linguistic regularities (Levy & Goldberg, 2014b) and no substantial or striking difference is observed between the two kinds of models (Levy, Goldberg, & Dagan, 2015; Mandera, Keuleers, & Brysbaert, 2017).

One last general aspect of DSMs needs to be highlighted before turning to the analysis of what kind of linguistic data can actually be extracted on the basis of the methodological implementations discussed above. Both window-based count models and word2vec predict models can be implemented taking into consideration the syntactic dependencies in context. DSMs models based relations of syntactic dependency (e.g.: Baroni & Lenci, 2010; Levy & Goldberg, 2014a; Padó & Lapata, 2007) represent a more linguistically informed version of window-based models,

which adopt a surface-oriented perspective. In contrast, dependency-based DSMs adopt a syntactic perspective on co-occurrence: "nearness" is defined by the presence of a syntactic relation between target and features (e.g. direct object, subject, adjectival modifier), which may also correspond to a path along several edges of a dependency graph. If syntactic relations are only used to determine co-occurrence contexts, we talk of dependency-filtered DSMs; if the type of relation is explicitly encoded in the context features (e.g. "subj_dog"), we talk of dependency-typed DSMs (Lapesa & Evert, 2017). It is still an open question whether dependency-based model, which encode more linguistic information and need to specify the type of dependency relations that are selected as contexts, provide a real advantage over surface-oriented models. For instance, it has been argued that syntax-based models tend to identify distributional neighbors that are taxonomically related, mainly co-hyponyms, whereas linear window-based models are more oriented towards identifying associative relations (Peirsman, Kris, & Speelman, 2007; Van de Cruys, 2008; cf. Acquaviva et al., forth). However, large-scale evaluations of dependency-based models show that properly tuned window-based DSMs still outperform the dependency-based models in most tasks (Lapesa & Evert, 2017). Hence, the question of whether syntactic information provides a real advantage over bag-of-word representations in DSMs is open and debated, and the answer to it highly depends on the semantic task on which models are tested (Kiela & Clark, 2014).

5.3.2. What DSMs do: similarity, relatedness, relationships, and senses

We should now have a more or less precise idea of what distributional models are, what are their main parameters, the differences between the various models, and of how they learn. But, what do they actually learn? What use can we make of distributional representations?

The first basic operation that we can do is to compute the similarity (usually by computing the cosine distance) between vectors. According to the distributional hypothesis, words that appear in the same linguistic contexts are similar in meanings. Hence, distributional representations of words that appear in similar contexts should be closer in a continuous semantic space defined by mutual proximity relations among lexical items. In the example (figure 20), *football* is similar in meaning to *soccer* because the context vector for football (the row corresponding to *football* in the word-word matrix) has a large numerical overlap with the context vector for *soccer*; many of the

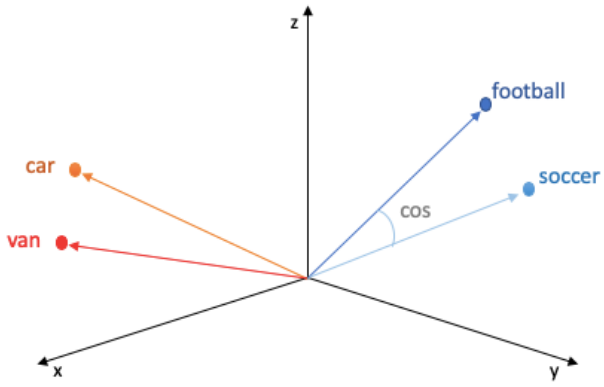


Figure 20. Example of distributional vectors

words surrounding instances of *football*—within a contextual window of a sentence—are the same as the words surrounding instances of *soccer*.

Besides this very intuitive measure of similarity, the notion of similarity itself is very complex and deserves an in-depth consideration. This intuitive notion of similarity tells us that words that are close in semantic space are, in some way, similar. However, words can be similar in many ways, in the sense that semantic similarity is a cover term for many different types of lexical relationships, e.g. synonymy, hypernymy, antonymy, meronym, locative relations, and topical and other non-classical relations. Furthermore, for pairs of words such as *doctor* and *hospital* or *car* and *driver* we would not say that they are semantically similar, but rather semantically (or, more specifically, thematically) related[151]. These two types of relations have very different properties. Yet they are barely distinguished by DSMs. In general, the semantic neighbors identified by DSMs have multifarious relations with the target, suggesting that DSMs provide a quite coarse-grained representation of lexical meaning.

Hence, on the one side, semantic relation discrimination is an important area of research in DS and dataset can be designed especially to task DSMs on this issue (Baroni & Lenci, 2011; Santus, Yung, Lenci, & Huang, 2015). On the other side, already ongoing research in DS has shown that specific parametrizations of DSMs - especially the manipulation of the context taken into consideration - biases models to capture specific kinds of relationships: for instance, it has been shown that semantic similarity is better identified by models based on strict definitions of the

[151] For a discussion of the notion of *thematic* cf. Chapter 4.

context definition; that on detecting semantic similarity syntactic approaches perform better than bag-of-word models, and small context windows are better than larger ones. For semantic relatedness, however, syntactic features and small context windows are at a clear disadvantage. Narrow-context models give high similarity for synonyms, hypernyms and co-hyponymies (Peirsman, 2008).

All these observations are indeed interesting but remain still too vague for the purposes of our discussion. Given this overall lack of specificity of the kind of semantic similarity/relatedness that is extracted by means of DSMs, what can we say about the representation of a semantically structured space? In summary, although DH is couched in terms of similarity, DSMs seem to actually be more biased toward the much vaguer notion of semantic relatedness. The outcome of DSMs still resembles a network of word associations, rather than a semantically structured space (Lenci, 2018). Even though this is a clear weakness of distributional approaches, some considerations can already be made about the kind of structure observed by means of distributional representations in a relationally-build semantic space.

As we have introduced in the previous section, one first distinction in classifying distributional models with regards to the kind of structural behavior they allow to extract is that between syntagmatic and paradigmatic models, i.e. between models that represent syntagmatic relations (co-occurrences) and models that represent paradigmatic relations (neighbors) (Sahlgren, 2006, 2008). Syntagmatically related words can be defined as words that co-occur in the same text region, from simple collocations, like “hermetically sealed”, to words that co-occur with a possibly even large number of words between them. The most relevant parameter in syntagmatic models is the size of the context region taken into consideration. Syntagmatic models collect text data in a word-by-document co-occurrence matrix in which the cells indicate the (normalized) frequency of occurrence of a word in a document. Paradigmatic models collect text data in a word-by-word co-occurrence matrix that is populated by counting how many times words occur together within the context window. The majority of the models we considered so far are of a paradigmatic kind: similarity of words in context (the cosine distance between vectors) is generally computed between neighbors. However, Lapesa and colleagues (2014) argue that paradigmatic models are not constrained to the individuation of only paradigmatic relationships as bag-of-words models perform well on both paradigmatic and syntagmatic relations if their parameters are properly tuned. In particular, their results show that size of the context window and the selected technique

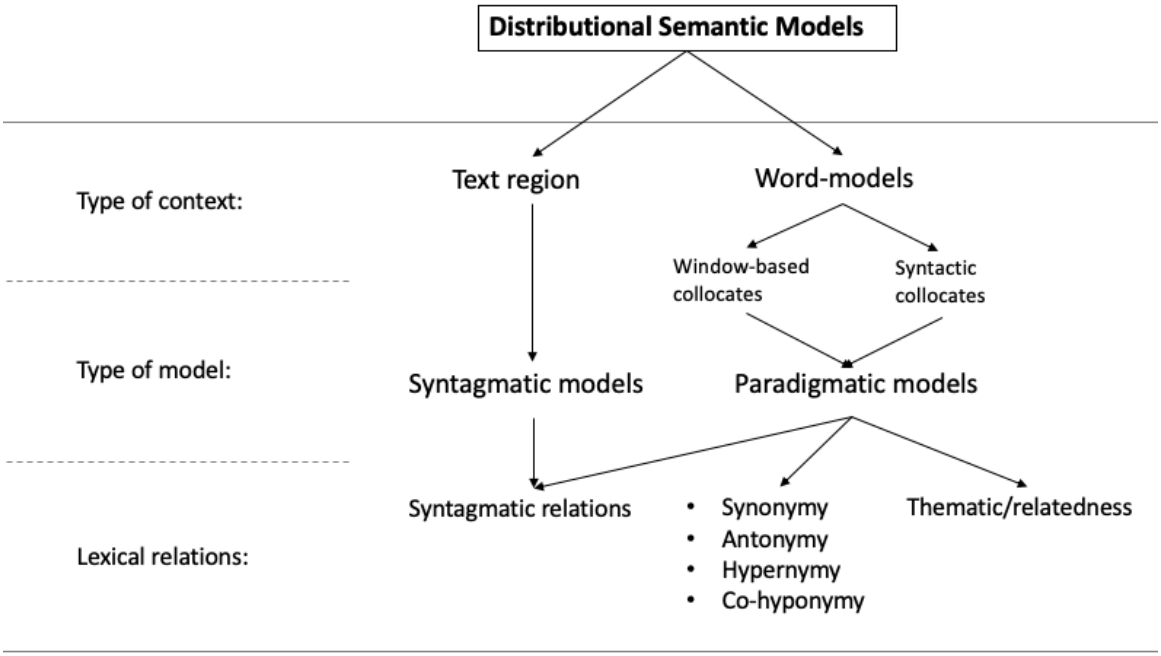


Figure 21. Types of DSMs based on context-type and schematic representation of the kinds of lexical relations identifiable with each model.

of dimensionality reduction play a key role in differentiating DSM performance on paradigmatic vs. syntagmatic relations.

Can we distinguish between different types of lexical relations within a distributionally constructed semantic space? The possibility of detecting synonymy is the most straightforward interpretation of distributional data (Curran, 2003; Landauer & Dumais, 1997). Interestingly, distributional representations allow to understand synonymy and synonymic relationships, as well as semantic similarity in general, as a matter of degree. Considering that distributional representations are highly context-sensitive and that they implicitly represent the different senses of a word given its linguistic context, it is extremely unlikely that two words, even if they are close synonyms, would appear in the exact same linguistic contexts and hence would have the exact same vector representation; "words are near-synonyms if they have nearly identical contextual representations" (Schütze, 1997, p. 88). Beside synonymy detection, which is indeed the most common linguistic benchmark for DSMs, most research in the area has focus on hypernyms identification (Lenci & Benotto, 2012; Santus, Lenci, Lu, & Schulte im Walde, 2014; Shwartz, Santus, & Schlechtweg, 2017). Antonymy also represents a significant challenge for DSMs as

synonyms and antonyms are often indistinguishable in terms of their degree of distributional similarity (Mohammad, Dorr, Hirst, & Turney, 2013) since both tend to occur in similar context (Miller & Charles, 1991). Current approaches to determining antonymy range from identifying contexts that are expected to be more discriminative of contrast (Santus, Lu, Lenci, & Huang, 2014; Turney, 2008) to using hybrid models in which DSMs are enriched with information extracted from other lexical resources.

Beside the identification of specific semantic relationships, distributional representations can be used to disambiguate between different word senses and to tackle onto the problem of polysemy.[152] Corpus-based distributional models are unable to model distinct word meanings as they conflate the contextual evidence of different meanings of a word into a single vector. However, distributional representations have now proven to be a reliable source on the basis of which we can identify different senses of ambiguous (polysemous) words when they are integrated into downstream applications. The general take is that we can induce word senses from raw text by clustering word co-occurrences.

The distributional hypothesis for word senses can be formulated as follows: "two occurrences of an ambiguous word belong to the same sense to the extent that their contextual representations are similar" (Schütze, 1997, p. 88). The analogy between the contextual hypotheses for words and senses is that both word senses and word occurrences are semantically similar to the extent that their contexts are semantically similar. A group of contextually similar word occurrences is a sense: two occurrences belong to the same sense if they have nearly identical or at least very similar contexts. Hence, a first set of approaches in word sense disambiguation within the framework of the distributional hypothesis is based on the notion of *context clustering*. Each

[152] Word Sense Disambiguation (WSD) is a complex field of research, in which data from distributional representation play one role in a much broader framework of computational techniques and research questions. In very general terms, WSD can be viewed as a classification task: word senses are the classes, and an automatic classification method is used to assign each occurrence of a word to one or more classes based on the evidence from the context and from external knowledge sources. The skeletal procedure of any WSD system can be summarized as follows: given a set of words (e.g., a sentence or a bag of words), a technique is applied which makes use of one or more sources of knowledge to associate the most appropriate senses with words in context. Knowledge sources can vary considerably from corpora (i.e., collections) of texts, either unlabeled or annotated with word senses, to more structured resources, such as machine-readable dictionaries, semantic networks, etc. Hence, distributional models are just one of the possible model-sources (unsupervised WSD) for the application of sense disambiguation techniques, together with supervised models and knowledge-based WSD. Unsupervised WSD are actually Word Sense Induction (WSI) techniques, in the sense that they aim at discovering senses automatically based on unlabeled corpora (for a general introduction and survey cf. Navigli, 2009, 2012; Pal & Saha, 2015).

occurrence of a target word in a corpus is represented as a context vector. The vectors are then clustered into groups, each identifying a sense of the target word.[153]

In recent years, multiple methods have been developed at the intersection of distributional semantics and word senses disambiguation techniques (and especially of word embeddings obtained by means of predict models) (e.g. Chen, Liu, & Sun, 2014). For example, Iacobacci and colleagues (2016) showed that a WSD system that makes use of word embeddings alone, if designed properly, can provide significant performance improvement over a state-of the-art WSD system not implemented on distributional representations. In a previous study (Iacobacci, Pilehvar, & Navigli, 2015), they proposed an approach to obtain continuous representations of individual word senses, referred to as sense embeddings. More generally, there is an increasing interest in the NLP literature in facing the conflating deficiency of distributional models, progressively shifting from a word to a sense representation able to capture linguistic phenomena such as polysemy (for a very exhaustive review of current methods and approaches cf. Camacho-Collados & Pilehvar, 2018; Navigli & Martelli, 2019).

In conclusion, even if, at the current stage of research, the outcome of DSMs does not look like a properly structured semantic space, the information recorded in the word vectors allow us, with properly adjusted computational methods, to tackle onto various kinds of lexical relationships as well as onto other structural/relational aspects of the meaning of words, such as the representation of the multiple senses that words convey in different contexts.[154]

---

[153] A context vector is built as the centroid (the normalized average) of the vectors of the words occurring in the target context, which can be seen as an approximation of its semantic context. These context vectors are second-order vectors, in that they do not directly represent the context at hand. Sense discrimination can be performed by grouping the context vectors of a target word using a clustering algorithm. On this line of research, Schütze (1998) proposed an algorithm, called context-group discrimination, which groups the occurrences of an ambiguous word into clusters of senses, based on the contextual similarity between occurrences.

[154] Two further points need to be mentioned in order to fully grasp the reach of DS in modeling linguistic data within the domain of linguistics. The approaches we have discussed so far are mainly concerned with lexical semantics and hence do not consider the fact that a central aspect of human semantic competence is the ability to compose lexical meanings to form the interpretation of a potentially unlimited number of complex linguistic expressions. To this regard, recent developments in DS have given rise to Compositional Distributional Semantics, a branch of DS interest in evaluating if and how DS can be integrated with (formal) compositional views of meaning. Similarly, research in compositional DS also deals with a broad range of phenomena at the intersection of formal and distributional semantics, such as meaning variation in context, selectional preferences, coercion, and thematic fit (for a review and relevant references cf. Lenci, 2018). Although recognizing the centrality of this new approach for the development of DS approaches, we will not extensively discuss them here as our whole argument is mainly about lexical semantics as well. However, it seems evident that these developments of DS in relation to compositionality bring DS approaches even closer to usage-based approaches to linguistic meaning.

5.4. Summary and (partial) conclusions

In this chapter, we have touched on many different topics in linguistics and we have reviewed a variety of approaches that, at least to us, seem to point in the same direction: linguistic meaning is something we construct by abstracting it from the uses we make of language. Despite sharing this fundamental intuition about the nature of meaning, the theories we discussed often come from different backgrounds and focus of slightly different aspect of these processes of meaning construction as abstraction from standardized uses. In this brief conclusion, we aim at summarizing the insights from the different perspectives we outlined above and at highlighting their commonalities and differences.

We could start by asking why we chose to start from cognitive linguistics, since the main line of our argument brought us to a somewhat different nature of meaning when compared to cognitivist claims. What is the role of paradigms of research such as constructionism in the development of our argument for a "language-specific" level of meaning? In various moments throughout this work, we have been arguing against a pure conceptualist view of meaning: from the very beginning we tried to point out some of the fundamental shortcomings of the idea that language simply labels concepts or portions of an already structured conceptual space. It is nonetheless clear, that concepts play a role in defining what linguistic meaning is, and hence, that the relationship between language and thought (e.g. how much of language is actually a map onto our conceptual structure) cannot be left unexplored and it is essentially related to the need of establishing what language is and how it operates at the cognitive level. Second-generation cognitive linguists had the great merit of recognizing (against generative approaches) that language is essentially grounded in non-language specific cognitive mechanisms and that it interacts and is integrated with the rest of our cognition. For instance, constructivist approaches make this point very clear: the way in which languages come into being involves cognitive processes such as categorization, chunking, and so on.

If for the processes involved in language formation and change this cognitive view seems more than plausible and it is receiving support from studies in both developmental psychology as well as in language evolution, understanding the content of language representations and where and how we get if from seems to be less straightforward. The identification of words with concepts has outstanding limitations: in the first place, the notion of concept itself is often extremely

underspecified and hence does not really favor us in shedding light onto how the mapping should work. In the second place, there is clear evidence showing that there are differences (and cross-linguistic difference make this point very clear) in the ways in which concepts (or cognitive abilities in general) and language may segment the space of experience. Constructionist approaches, and especially Langacker's Cognitive Grammar, have the advantage of shifting the focus from the notion of concept to that of conceptualization, from a static to a dynamic view of meaning, from a purely internalist view to a perspective deeply committed to understanding how we engage with the world and how language happens in that interaction. Constructivist approaches are indeed cognitive, in the sense that meaning is viewed as conceptual, but they shift towards a view in which meaning is not something that is just encoded in language or in our conceptual system, but it is instead something we build in interaction. As we have seen, one of the main tenants of constructivist approaches in that of being essentially usage-based. The mental grammar of speakers is shaped by use and development, i.e. by the repeated exposure to specific utterances, and domain-general cognitive processes such as categorization and cross modal associations play a crucial role in the entrenchment of constructions. Experience with language creates and impacts the cognitive representation for language. Hence, language uses are at the very core of the process that allows us to build meaning and to acquire a language in the first place. We learn and construct meaning from our experience with language.

In our view, one of the limitations of constructivists approaches is that they mostly focus on grammar and on the construction of grammar, whereas the meaning of constructions (form-meaning pairings, i.e. the linguistic symbols) is still thought as being something conceptual and entirely dependent on our embodied and grounded experience. Our argument can be viewed as a first (extremely sketched) attempt to draw a line of conjunction between usage-based theories of grammar and a renewed usage-based perspective on (lexical) meaning. As we have already admitted, within this work we were at best able to set a possible path towards the integration of the two (i.e. to actually build upon constructivist notions to get to a usage-based theory of meaning) and we did not even scratch the surface of the argument. In our view, the argument we proposed is somewhat preliminary to a full-fledge account of a usage-based theory of meaning. Both theoretical reflections and evidence from a variety of psycholinguistic experiments point in the direction of recognizing that viewing language as a direct window on our conceptual structure is at best partial. Hence, our proposal (developed in this and the next chapter) aims at delineating

why and how we should focus on a "language-specific" level of linguistic meaning. In very general terms, while the conceptualist identification of meaning appears to be viable for a certain class of referents, it is globally far less successful than usually assumed. As a result, research oriented towards revealing the conceptual content underlying the entire lexicon may have less to tell us about language structure than typically believed. Concepts (non-linguistic categories and knowledge) may only play a partial role in linguistic semantic. Construction grammar poses the seeds for theorizing a "language-specific" level of linguistic meaning within the cognitive tradition.

Even though the term "usage-based" was coined within the second cognitive revolution, constructionist approaches are indeed not the first to recognize that the uses we make of language and the contexts in which we use it are fundamental in determining language, meaning, and its structure. The Saussurian perspective, as we have been presenting it throughout this work, is a clear example of it. The arguments proposed by Firth and the London School point in the direction of grounding the descriptive study of language in the context of situation, of which the text comes out as a possible first approximation, and that the study of meaning needs to be carried out at multiple levels of analysis. The discussion of some of the main positions in Firth's work allowed us to further ground the view of the systematicity of language as arising from the uses we make of it and to first glance at the idea that meaning needs to be unpacked in its multiple components if we ought to give a comprehensive account of it. Needless to say, Firth's notion of collocation and Harris's operationalization of the notion of distribution are the long-claimed grounds of distributional semantic approaches. Our discussion of DS moved two ways: on the one side, we moved in the direction of explicitly grounding DS in the Saussurian soil (the historical review essentially aims at this); on the other side, we tried to show how DS allows us to tackle onto the "language-specific" relational aspects of linguistic meaning by analyzing the kind of meaning that distributional representations encode and the kind of relationships they allow to represent.

Indeed, one of the main questions and concerns raised against DS is what kind of meaning it actually allows to tap onto. Can meaning actually be described at a linguistic level? Does this kind of analysis bring us any closer in understanding the nature and structure of linguistic meaning? Although we recognize that these questions are far from being actually resolved, we argued that the answer to these questions may indeed be *yes*. DS can be essentially thought of as a way of operationalizing the notion of system and to shed light on (at least some) of the relational

aspects of meaning. We have seen how both the Firthian and the Saussurian tradition (although with substantial differences) highlight the need to view meaning as (at least in part) a linguistic fact: in our terms, the "language-specific" level of linguistic meaning can be identified in the relations that the linguistic units entail within the language system, which is always the outcome of the stability provided by the system itself and the variability of the individual uses in the acts of speech. Distributional semantics is in line with these theoretical assumptions and allows us to find structure in uses providing the methods by means of which we can describe the relationships internal to the language system.

In conclusion, in this chapter we have mostly embraced a linguistic perspective, i.e. we have tried to outline a coherent view that would allow us to encompass and explain some fundamental facts about language. Even from a strictly linguistic perspective, our work is far from being done and many points still have to be addressed or fully developed, both within DS itself as well as from our perspective of taking DS as a means by which we operationalize the relational system of meaning. Moreover, as our incipit to this chapter shows, we cannot claim to have described meaning if we ignore the cognitive side of it (which is indeed the most debated aspect of current approaches on meaning). Moving forward, we will try to outline the role that distributional semantics plays in explaining meaning in cognition and how it can be integrated in a unitary framework in the study of meaning that aims at giving a comprehensive account of language in language and of language in cognition.

Chapter 6

Distributional semantics in cognition: what do we learn from language?

6.1. Distributional semantics and the study of language in cognition

In the previous chapter we have limited ourselves to the discussion of DS approaches from the point of view of linguistics: DS approaches have their roots in the Saussurian view of the language system, and to some extent in Harris's structural methodology, enlarged within the framework of Firth's notion of context of situations, exemplified in the study of texts, and of the centrality of the notion of collocation in the study of linguistic meaning. From a theoretical perspective, DS approaches allow us to represent the structural-relational aspects of the language system within the framework of a usage-based account of linguistic meaning. Hence, DSMs grant us to isolate what we have defined as the "language-specific" level of linguistic meaning by means of mathematical-statistical techniques. As we have pointed out in the previous section, the question of what specific type of structural relationships can be identified by DSMs is still open and the results highly depend on the type of model and on how its various parameters are set. Nonetheless, even if there is a long way to go in experimental research before actually being able to obtain a structural network for distributional representations, and it could even be the case that this might not at all be possible, new findings, research questions, and theoretical observations all lead towards corroborating the idea that distributional representations actually allow us to tackle onto the relational aspects of linguistic meaning.

In addition to the perspective from linguistics, DS approaches have seen and still see a great interest in psycholinguistic approaches to meaning. As we have argued multiple times already, the notion of meaning in psycholinguistics has a much broader definition compared to more specifically linguistic definitions. It seems undeniable that, if taken form a psychological perspective, the notion of meaning is indeed very complex and likely to encompass a wide variety of bodily, conceptual, linguistic, social, cultural aspects and so on, which do not really seem to be easily isolable one from the other in the intricate maze of our encyclopedic knowledge. Although

the linguistic context appears as one of the ingredients of human conceptualization, the emphasis of cognitive semantics is on an intrinsically embodied conceptual representation of aspects of the world, grounded in action and perception systems. The distributional constraints to which linguistic constructions obey are intended to receive a functional explanation in terms of the principles governing the processes the underlie our conceptualization of the world. In our discussion up to this point, we have mostly focused on how distributional semantics can account for many facts of language from a linguistic perspective. In this section, we will focus our discussion on if and how distributional semantics can shed light to (at least some aspects) of how language is learned and stored at a cognitive level.[155] The fundamental question underlying the interest of cognitive approaches to distributional semantics concerns the possibility of a causal role of the contexts in which words are produced and observed in determining the structure and organization of semantic representations at the cognitive level.

6.1.1. The "strong" distributional hypothesis revisited: an outline

The "strong" version of the distributional hypothesis offers a fairly straightforward hypothesis: "repeated encounters with words in different linguistic contexts eventually lead to the formation of a contextual representation as an abstract characterization of the most significant contexts with which the word is used" (Lenci, 2008, p. 16). The distributional behavior of a word in the linguistic context can be taken to explain its semantic content at the cognitive level. Before turning to the evaluation of this hypothesis in light of the psycholinguistic data available in the literature as well as the most common psychological benchmarks of DSMs, we will briefly introduce how this hypothesis came into being in the first place.

Drawing from Harris's definition of distribution and from Firth's famous claim that "we shall know a word by the company it keeps" (Firth, 1957), Miller and Charles (1991) argue for a "usage-based" characterization of semantic representations:

[155] In this section we will not address one of the aspects that has been central to our discussion of the mental lexicon, i.e. language processing. Hence, our argument here will be mostly focused on the level of the mental content rather than on a model of how meaning is processed at the cognitive level. At the same time, the literature reviewed seldom mentions this topic. Predict models are based on a distributional idea of representation and on a neural kind of architecture, both in line with connectionist approaches to language and cognition. However, we are a long way from being able to properly address the cognitive plausibility of (for example) predict models in terms of how human language processing actually unfolds (for a review and interesting insights cf. Mandera, Keuleers, & Brysbaert, 2017)

> What people know when they know a word is not how to recite its dictionary definition – they know how to use it (when to produce it and how to understand it) in everyday discourse [...]. Knowing how to use words is a basic component of knowing a language, and how that component is acquired is a central question for linguists and cognitive psychologists alike. The search for an answer can begin with the cogent assumption that people learn how to use words by observing how words are used. And because words are used together in phrases and sentences, this starting assumption directs attention immediately to the importance of context (Miller & Charles, 1991, p. 4).

This leads to the idea that people can learn how words are used on the basis of information provided by purely linguistic context, constructing "contextual representations" as some form of abstraction or generalization derived from the contexts that have been encountered. These contextual representations include all "syntactic, semantic, pragmatic, and stylistic conditions governing the use of that word" (Miller & Charles, 1991, p. 8).

Hence, the strong version of DS is a cognitive assumption about the form and origin of semantic representations, assigning a specific causal role to word distributions in the formation of the (cognitive) semantic representation for that word. Given these theoretical premises, the strengths of DS as an explanatory model of meaning needs to be evaluated on the grounds of the semantic facts that it is actually able to explain. As we will see, there is a growing body of evidence that shows that the probabilistic analysis of linguistic distributions has a significant role for language organization and for its dynamics. On the one side, we will endorse a view that sees linguistic distributions as one of the ingredients in the formation of our mental representations, and not as *the* only one. This perspective is becoming more and more central in those approaches that seek to explain the formation and nature of meaning in cognition, especially at the intersection of embodied and distributional perspective, both in the psycholinguistic literature as well as in approaches interested in modeling cognitive processes. On the other side, based on the argument we provided in support of the idea that DS allows us to tackle onto the "language-specific" level of linguistic meaning, we will argue that the structure observable in distributionally constructed semantic spaces cannot be thought of, at least from a theoretical standpoint, as a direct representation of the structure of meaning in the mind, which ultimately seems to depend on a wider variety of environmental, physical, and cultural factors. Similarly, endorsing a view that sees distributional representations as one ingredient in the formation of concepts, hence a view close to the "strong" distributional hypothesis, does not imply neither that distributional models of linguistic meaning should be viewed as models of cognition, nor that the mechanisms underlying

these models (e.g. the algorithms by which them learn or the cosine measure of similarity) also guide language acquisition or language processing, such as storage or retrieval. [156] Let us first take a step back and briefly review the most common cognitive benchmarks of distributional approaches to meaning, i.e. how distributional approaches can help us understand the format and structure of semantic memory, of language processing in cognition, as well as how people judge, evaluate, make attributions about objects and concepts, and so on.

6.1.2. The limits of testing DSMs on cognitive benchmarks

Ever since Landauer and Dumais (1997) demonstrated that distributional semantic representations could be used to make predictions about human performance in semantic tasks, numerous researchers have used measures of (dis)similarity between word vectors - e.g. cosine similarity, euclidean distance, correlation - for that purpose. DSMs are generally evaluated for their correlation with different kinds of semantic-related tasks and, hence, for how well they can predict language-related behavioral data. DSMs' performances are contrasted with various sets of human data in order to evaluate their efficiency in producing human-like linguistic behaviors. Many studies have focused on the performance of a single model in a variety of tasks or on a comparison between different models on the same set of tasks (e.g.: Mandera et al., 2017; Pereira, Gershman, Ritter, & Botvinick, 2016). In general, DSMs obtain very good results in a wide range of semantic tasks: multiple-choice synonym detection tasks like the one used in the Test of English as a Foreign Language (TOEFL) (Bullinaria & Levy, 2012; Landauer & Dumais, 1997), similarity ratings as collected in data sets such as SimLex999 (F. Hill, Reichart, & Korhonen, 2015),

[156] One interesting exception is the very general mechanism that seems to underlie both distributional models and human language acquisition, as well as other cognitive mechanisms; i.e. distributional semantic spaces are built on computational models (including neural networks) that use domain-independent learning algorithms recording the distributional statistics in the linguistic input. A long tradition in both psychology and psycholinguistics has been interested in evaluating the role and mechanisms of implicit learning and statistical learning (and especially in the cross-pollination of the two, giving rise to the idea of "implicit statistical learning": Christiansen, 2018), i.e. of how learners can implicitly use distributional properties of the input to discover patterned regularities in their environment. Mechanisms such as implicit and statistical learning seem to underlie both distributional methods, especially those in line with connectionist approaches, as well as human language acquisition. We will not further develop this point here as, although playing a central role, it is more a matter of understanding and modeling general aspects of human cognition rather than of understanding how distributional representations of the "language-specific" structure of language can be integrated in our conceptual knowledge and what we actually learn from our experience with language.

categorizing nouns (Baroni & Lenci, 2010; Riordan & Jones, 2011), semantic priming (Günther, Dudschig, & Kaup, 2016; Jones, Kintsch, & Mewhort, 2006; Mandera et al., 2017), and analogy detection (Mikolov, Chen, et al., 2013).[157]

The underlying assumption that seems to implicitly justify this evaluative methodology is the idea that semantic similarity as abstracted from language data corresponds (or at least has a high correspondence) with word similarity in the mental lexicon, and that this notion of semantic similarity could account for how words are represented in memory and for how they are semantically associated. Hence, an interesting cognitive benchmark for DSMs becomes how they can simulate human word association data. This direct link between distributional similarity and word associations is not historically unjustified: the primary and most specific goal of Miller and Charles (1991) was to use contextual representations to provide an empirical characterization of semantic similarity, a central notion in psycholinguistic research. In their view, semantic similarity (as a cognitive notion) is to be treated as a dependent variable of the contexts in which words are used, i.e. as a function of their contextual representations. As Miller and Charles also point out, the mechanisms leading to the formation of contextual semantic representations can be viewed as a specific instance of general cognitive associative mechanisms recording the statistical co-occurrence between stimuli. Furthermore, the associations made between words are interpreted as the result of the relative distribution between stimulus and response in linguistic contexts, and the stimulus-response association strength is related to the distributional similarity between the two words (J. J. Jenkins, 1954; cf. Lenci, 2008 for a discussion).

This claim has a strong neo-behaviorist taste. Associationism is often taken as one of the fundamental aspects of behaviorist approaches in psychology, in which word associations are taken as evidence to analyze the organization of language and, according to the view described above, the distribution of words in context would suffice to shed light on the kind of associations we build between words (for a discussion of the notion of association in psychology and its operationalization in word association norms cf. McRae & Jones, 2013). On the one side, the brain is indeed, among other things, a very powerful associative device and co-variation can be regarded as a general property of human experience and neural processes that reaches beyond word co-

[157] Many more examples could be provided of studies that both speak for the ability of distributional representations to account for psycholinguistic data (for a review cf. Mandera et al., 2017; Pereira, Gershman, Ritter, & Botvinick, 2016) as well as that highlight the limitations of distributional data to this regard (e.g. Griffiths, Steyvers, & Tenenbaum, 2007; Nematzadeh, Meylan, & Griffiths, 2017).

occurrence (de Vega, Graesser, & Glenberg, 2008). On the other side, it is now widely recognized that associative mechanisms do not suffice to explain cognitive dynamics, and this seems to pose important questions to the validity of distributional information in the formation of concepts.

Whether or not associative mechanisms are at the basis of our cognitive abilities and of how we construct and store information, the direct link between words that co-occur and words that are associated seems to be ill posed. In the first place, the notion of semantic similarity is claimed to have a central and widespread role in cognition above and beyond the formation of associative pairs; we group together similar things in a way that cannot be accounted for by mere associations, and this is also seems to be the main source of categorization and concept formation processes. In the second place, posing a role for distributional information in the formation of associations between words does not necessarily imply that other aspects of perception or cognition may also play be involved and, hence, that the formation of word associations may not entirely depend on direct links of words co-occurring or sharing the same context. In the third place, free association data is the product of a retrieval process operating on a semantic representation, and we should not mistake the similarity metric (cosine) for an appropriate process model of the association task; cosine is not what people do with a memory representation.

Recently, Griffiths, Steyvers, and Tenenbaum (2007) demonstrated that spatial models are unable to simulate human free association data due to the constraints placed upon them by metric axioms which appear to be violated in association norms (cf. also Nematzadeh, Meylan, & Griffiths, 2017).[158] Word association norms contain a significant number of asymmetric associations: for example, the probability of generating *baby* as a response to *stork* as a cue is much greater than the reverse. In response to such critiques and limitations of DSM data to account for word associations, Jones and colleagues argue that "semantic space models are models of memory structure; the structural model should not be expected to simulate a complex behavior like memory retrieval without the benefit of a process account to explain how the memory structure is

[158] Griffiths et al. (2007) note that probabilistic representations are not subject to the same metric restrictions as spatial representations, and they provide an elegant demonstration of how *topic models* can naturally account for the qualitative nature of violations in asymmetry and the triangle inequality where, for example, LSA cannot. Griffiths et al. further demonstrate that while LSA (based on a thresholded cosine) cannot reproduce the scale-free and small-world network structure seen in word association norms, this structure naturally emerges in a *topic model*. Utsumi (2015) has revisited the first Steyvers and Tenenbaum (2005) work and demonstrated that while scale-free and small-word structure is unobtainable by LSA, several other variants of the model, all spatial models, naturally produce the correct structure from association norms.

used in a particular task. While the cosine between two word vectors is often used as a measure of their semantic similarity, it is a measure of the similarity of memory structures rather than an appropriate process model of the task" (Jones, Gruenenfelder, & Recchia, 2018, p. 55). In the paper, the authors show the ability of spatial semantic models to simulate association data when they are fused with a simple Luce (1959) choice rule to simulate the process of selecting a response in free association, and hence that spatial models can produce the structure observed in free association data provided that they have a plausible process model to simulate the association task.

At a more theoretical level, according to our understanding of what kind of semantics is distributional semantics (i.e., DS approaches allow us to represent the "language-specific" level of meaning) and according to the theoretical background we have been discussing so far, the distributional semantic space and the distributional similarity between words should not be thought of as simply mirroring our mental content or as a direct window onto its organization. If we take the semantic space designed by means of free word association data as a (possible) representation of the organization of our mental contents, we should not expect distributional information, which extract and represent the organization and structure of the language system, to simply reflect it, as the mental structure of our encyclopedic knowledge derives from and comprises information that is not only linguistic or related to the organization of the language system, but also grounded in our physical, emotional, and cultural experiences.

Our hypothesis is that distributional data and word association data represent slightly different (but intertwined) aspects of our knowledge about language and of our general knowledge. Distributional networks are thought of as tackling onto the structural aspects of the language system and allow, first and foremost, to represent the relationships between words that have been standardized in language use. In turn, word association networks can be thought of as a representation of a much broader notion of meaning and, consequently, of a different structure of our knowledge, more broadly grounded in the variety of human experience. However, as we have seen, the two aspects (language in its structural organization and meaning in cognition) are not completely separated in the sense that the way in which we do meaning (i.e. we communicate, but also the way in which we constantly reshape our mental contents as well as the pressure that idiosyncratic uses pose on the language system) can be defined as the constant negotiation between these two aspects. In this sense, part of our mental structure for meaning is defined by the "language-specific" linguistic structure and vice versa, but the two do not seem to be completely

overlapping (see Box 2). The distributional "language-specific" structure of language is one of the aspects that contributes in shaping our knowledge about the world both in development as well as in our adult life. Hence, one interesting line of research is that of understanding how we integrate the structural knowledge of language, that we encounter in communication and in exposure to language, with the rest of the physical, emotional, cultural inputs. The main point being that we learn something from language, the question is *what* and *how* we learn it.

6.1.3. Bits and pieces of the *how*: semantic memory in the distributional framework

Some suggestions about the *how* come from the literature interested in the study and modeling of memory for semantic knowledge, from which we can draw interesting insights about the format of our mental representations. In recent research paradigms, distributional information has been highlighted as an important factor in shaping our memory for (broadly understood) semantic knowledge. This is a fundamental aspect to take into consideration when discussing and elaborating models of semantic knowledge, i.e. of how knowledge is represented and used.

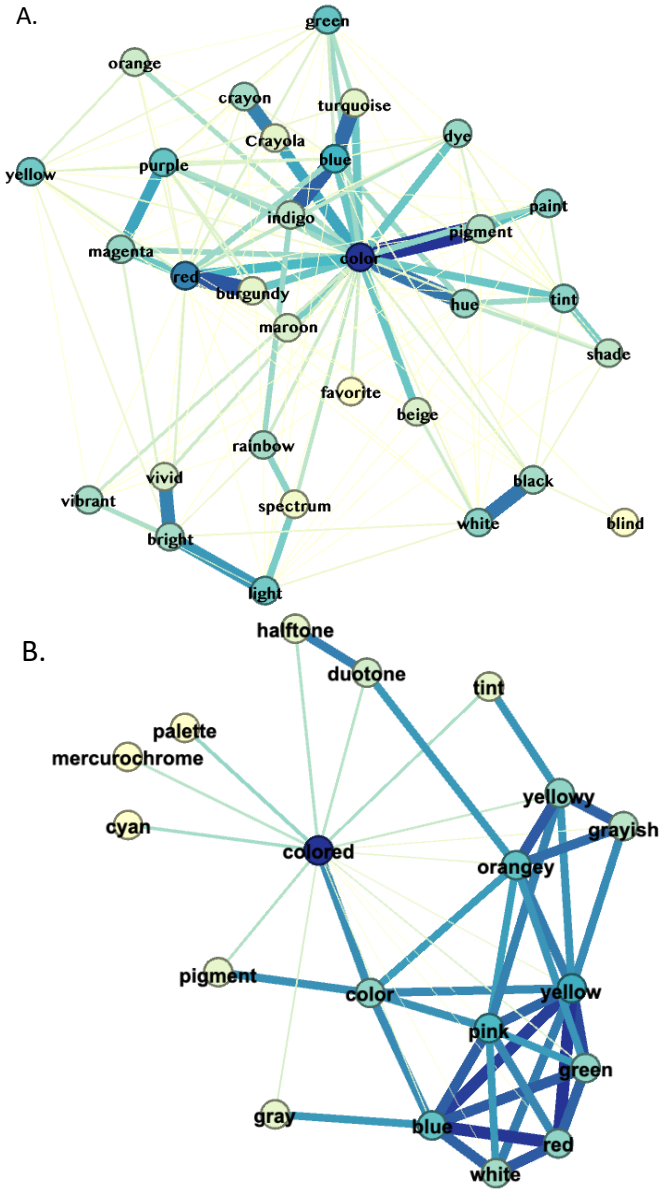
In the eyes of researchers interested in how we create and organize memory, the common theme, unifying all distributional approaches and models, is that they hypothesize a formal cognitive mechanism to learn semantics from repeated episodic experience in the linguistic environment. In this sense, distributional models of semantic memory are in line with episodic theory of memory formation. Computational models of distributional semantics differ in their learning mechanisms, but they all have the same overall goal of formalizing the construction of semantic representations from statistical redundancies in language. Hence, distributional models can be taken to provide insights in the mechanism that underlies the construction of semantic representations from statistical experience with language and, thus, to overcome some of the limitations of previous models (such as Quillian's) that did not specify how the network of representations could have been learnt (McRae & Jones, 2013).

In general, distributional models have been successfully used to better explore the statistical structure of the environment and to understand the mechanisms that may be used to construct semantic representations. Connectionist models are an excellent complement to distributional models, elucidating our understanding of semantic processing, and how semantic structure interacts with other cognitive systems and tasks (Jones, Willits, & Dennis, 2015).

Whether or not the mechanism (operationalized in the algorithm) underlying the functioning of DSMs actually mimics our cognitive processes, it is now widely recognized that linguistic information extracted from exposure to language is fundamental in how we organize and structure our knowledge, or at least our knowledge about language. As we have seen, the "strong" distributional hypothesis is essentially a hypothesis regarding how we acquire and construct meaning, and suggests possible mechanisms by which we accomplish that. There is a growing body of research suggesting that distributional information plays a more powerful role than previously thought in a number of aspects of language processing. The exploitation of statistical regularities in the linguistic environment has been put forward to explain how language learners accomplish tasks from segmenting speech (e.g. Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Toro, Sinnett, & Soto-Faraco, 2005) to bootstrapping syntactic knowledge (e.g. Gerken, Wilson, & Lewis, 2005; Monaghan & Christiansen, 2008; Reali, Christiansen, & Monaghan, 2003) and word meaning. Furthermore, distributional information seems to be fundamental not only for constructing our linguistic content and abilities, but also for shaping our conceptual content more generally.

Box 2. Free association and distributional representations

At the current stage of our research (but also of the research in general), this comes out as almost pure theoretical speculation and we currently have no data to back our position up. However, the problem of what kind of meaning is address in the two kind of approaches to the structure of language is generally underspecified and the research of the topic is still not offering a precise answer or way to investigate this point. At least in principle, we are offering a possible line of research in which the comparison between the two kind of data should be thought of as a comparison between structures representing different, not completely overlapping, portions of our knowledge. In general, studies have tried to analyze where and how the two kinds of structure can converge (for examples using the network approch cf. De Deyne et al., 2015; Utsumi, 2015). It seems to us that an attempt to explore the differences between the two structures and of how the might interact is theoretically justifiable and could offer important insights into how we construct and use meaning both at a cognitive as well as a linguistic level.



In this box we want to briefly report some ongoing research material and poke around with some merely speculative and preliminary observations, as a visual representation of what we have been arguing in the main body of the chapter could help visualizing our argument.

An example: the semantic space of colors

(A represents a network based on forward free word associations; B represents a network based on neighbors extracted from a CBOW distributional model)

Even if in this visual representation there seems to be a clear difference in the way the semantic space is constructed (i.e. there are apparent differences in the way the two graphs cluster), however, nothing can be said to this regard in terms of network analysis given that the networks were constructed for mere exploratory and visualizing purposes. It is nonetheless evident that, apart from the set of co-hyponyms for the word color (which bear the strongest connections – represented in the network edges) not many more overlaps can be spotted between the word associated to the word color by participants and its closest neighbors. This observation does not tell us anything in a definitive way, but it suggests that deeper investigations and comparisons could led us to interesting insight about the structure of our knowledge.

6.1.4. Introducing the question about *what* we learn

The question about *what* we learn from linguistic experience and *from what kind of information* we learn is at least as complicated as the question about the process of formation and the format of our representations. Understanding what we learn intersects the question about the content of our mental representation and attempts to address a much broader question about to what kind of experience we owe what we know. To what extent does our (broadly speaking) semantic knowledge depend on experiences with language? To what kind of linguistic experience do we owe such knowledge?

Before turning to an in-depth discussion of what we can learn from linguistic experience and of how linguistic knowledge can be integrated with the rest of our knowledge, allow us some final remarks. In 2011, IBM's Watson computer handily beat Ken Jennings and Brad Rutter, the world's top two Jeopardy players. Watson was able to benefit from fast retrieval algorithms and powerful inferential mechanisms. But while Kennings and Rutter had a lifetime of experiences interacting with the world on which they could draw, the source of Watson's knowledge came entirely from language (Ferrucci et al., 2010). The language that Watson ingested was, of course, generated by humans with direct experiences perceiving and interacting with the world. Still, that Watson was able to win relying entirely on the linguistic transmission of this knowledge, speaks to the power of language to encode and transmit such knowledge.[159] The information required to answer some questions, such as what Russian counts wrote "The Hedgehog and the Fox" is likely to have been learned from language by all the contestants. But faced with the clue "A piece of wood from a tree, or to puncture with something pointed", Jennings and Rutter could draw on their direct experiences with trees and sharp things. Watson had no such experiences. Yet it was Watson who answered that $600 question: "stick".

[159] A further point of recent discussions is that DS representations encode human-like biases and high-level judgments, suggesting that there is more encoded in language uses than previously thought. For example, DS models have recently been shown to encode the types of associative biases responsible for social stereotypes and prejudices (such as gender, race, and age biases) (e.g. Bhatia, 2017; Caliskan, Bryson, & Narayanan, 2017). The fact that DS representations encode human-like associations has one intriguing implication: by training DS representations on different corpora we can infer the associations that would be possessed by groups of people differentially exposed to (or responsible for producing) those corpora, and by doing so, infer differences in high- level judgments across these groups. Drawing on this insight, recent applications of DS models have attempted to study differences in social, political, and moral associations pertaining to media bias and social media structure (e.g. Bhatia, Goodwin, & Walasek, 2018) and to study changing gender, class, and ethnic associations over time (e.g. Garg, Schiebinger, Jurafsky, & Zou, 2018).

Watson's success does not tell us how much humans learn from language, but it does suggest that it might indeed be a lot more than previously thought. For example, when asked to list the features of an apple, sighted people tend to prominently mention the colors red, yellow, and green. Despite lacking all direct experience with the color of apples, blind people behave similarly, listing the same colors in the same order (though with lower frequency) (Lenci, Baroni, Cazzolli, & Marotta, 2013). For people born blind, the only source of such knowledge is language. Nonetheless, differences have also been found: congenitally blind subjects produced significantly less color terms when describing concrete objects than sighted control subjects (Lenci, Baroni, & Cazzolli, 2013). These contrasting evidence shows that, on the one hand, distributional information is rich enough to allow the organization of the color space to be derived from the linguistic input, while, on the other hand, the lack of direct perceptual experience may result in critical differences in the role and use of color information. What these examples seem to tell us is that our knowledge about the world can be viewed as the product of many kinds of experiences combined, in which language plays a specific (even if not easy – or maybe even possible - to delimit) role. The content of our mental representations (our knowledge about the world) arises from the integration of information we get from these very different sources or modalities of representation.

The role of linguistic and perceptual information as sources of semantic representation is still a puzzle with many missing pieces. New technologies that enable new experiments and data collected via different methods, should be considered as the methodological backbone of contemporary research in lexical semantics, and as maybe the most reliable way to fill these gaps. Experientially-based approaches to lexical semantics can provide evidence about how word meanings are construed, to what extent they are conventionalized, and how much they are influenced by perception and cognition or by cultural diversity and different typological properties, as well as the other way around to investigate what aspects of the already structured linguistic experience shape our mental representations.

6.2. To what do we owe what we know: multimodality goes both ways

In the previous sections we have mostly outlined and discussed the strengths of distributional approaches in representing word meaning from a relational and structural perspective as well as in accounting for a wide variety of psychological and behavioral linguistic data. To this last point, as we have already mentioned, the extent to which distributional data might explain cognitive processes involving language and semantic representations is highly debated, and one of the most relevant points of disagreement is, again, the notion of meaning itself. In fact, as soon as we step out of the linguistic description of meaning and we start wondering in the realm of cognitive explanations, arguing that meaning is "purely" linguistic is, at best, partial. As we have seen multiple times already, one of the fundamental stands of cognitive semantics is that meaning is, first and foremost, conceptual. We do not ought to deny this. Hence, how can these two aspects of meaning be kept together? How can we argue for distributional representations to tackle onto linguistic meaning and, at the same time, endorse a conceptual view of semantics? What is missing in distributional representations to be accounted for as full-fledged representations of meaning, given that they already allow us to reproduce and model many aspects of our semantic behavior? In this last section, we will try to address these points at the best of our possibilities and to propose a view that would allow us to entangle a distributional and a cognitive perspective on meaning in language and in the mind.

6.2.1. The state-of-the-art of our proposal

Before outlining the last fundamental steps of our argument, we ought to point out a few outstanding limitations. In the first place, our proposal is mostly theoretical: we draw from intuitions in linguistics and in psycholinguistics to explain data from empirical studies in psychology and psycholinguistic. This means that, at the current stage of our research, there are yet no empirical data specifically addressing (and hence corroborating) our proposal. In this sense, we think of this work as a new possible line of inquiry that could lead (and indeed leads) to interesting experiments and applications. In the second place, our proposal is not, for the most part, a complete novelty. Much of the research we will review in this section, and that is the basis for the construction of our argument, already lays down the majority of our building blocks. In a sense,

we are adding our interest for the structural and relational organization of linguistic meaning on top of an already flourishing paradigm of research. Our aim is to point out how the integration of the notion of language as a relational system with the current proposals could lead to possibly interesting insights in the study of meaning, cognition, and the interrelation of language and thought. In the third place, many of the main points of our argument have already been addressed in this and the previous chapters. Hence, our attempt here is to set out a coherent and comprehensive view of the positions scattered throughout our work.

In brief, our argument unfolds as follows: if we ought to account for meaning in its complexity, distributional models have indeed some limitations. Above all, distributional representations of meaning do not account for referentiality and are generally disembodied and ungrounded. Both aspects are fundamental for a comprehensive (and especially cognitive) theory of meaning; it is indeed true that we are able to denote present as well as absent objects and situations in the outside world and that a good part of what we call meaning is constituted by means of being grounded in action and perception. Hence, distributional approaches to meaning seem to suffer from the same downsides of the majority of symbolic accounts of meaning and cognition. In the first place, we will discuss in what sense is DS a symbolic account of meaning. In the second place, we will argue that, in line with what we have been outlining in the previous chapter, distributional representations allow us to tackle onto one specific aspect of meaning, i.e. the "language-specific" level of the relational and systemic organization of meaning in language, the web of pertinences that are standardized in the language system. Hence, DS does not impose itself as a comprehensive account of meaning and linguistic meaning is, already in itself, a multifaceted entity and does not simply mirror the structure of the external world or the structure of our conceptual contents. Moreover, the structure of language is, at the same time, constrained by the joints of nature as we perceive them and by the rest of our experience with the world. The way in which we negotiate meaning between our general knowledge of the world and the boundaries set by the language system implies that much of what is relational and structural actually becomes one of the ingredients in the formation of our concepts as well as that, in the opposite direction, the individual (and often idiosyncratic) content of our concepts is always injected in the language system, constantly reshaping it within (and even sometimes beyond) its limits. Therefore, meaning is always grounded and symbolic, cognitive and linguistics.

The key foundation of our argument is that much of what we know we have learnt from experience. The content of our mental representations is grounded in experience and one of the experiences we are always exposes to is the experience with language. Representations are multimodal, and one of their modalities is the linguistic one. The way in which we construct meaning is by integrating these various sources of knowledge. We will review some of the main arguments in favor of this view. Consequently, distributional representations allow to represent, surely with limitations, one central aspect of meaning in cognition. Distributional representations encode information that is different from that of perception, and the two are complementary. However, this is not an intrinsic limitation of distributional models per se, which have proven to efficiently support the integration of perceptual data. Multimodal and cross-modal distributional models seem to be able to address both the grounding problem as well as the problem of referentiality.

If framed within this perspective, it is now almost widely accepted that language plays a role in shaping our conceptual representations, in the sense that language is implicated in the formation of knowledge. As we will see, the inputs from language can be very rich and informative in a lot of ways. IBM's Watson computer and its skills are a clear example to this point. The main questions that remain open is concerned with what about language we acquire from linguistic experience and how we integrate it with the rest of our knowledge. Our hypothesis is that we learn from language its systemic organization, which shapes the way in which we organize and make pertinent our knowledge of the world. As we have already seen, the language system, which is integrated with and shapes our conceptual representation, can be operationalized and represented by means of distributional semantic approaches. In the first place, we will review some of the most relevant literature regarding *what* we learn from language (less attention will be here devoted to *how* we learn from language[160]). In the second place, we will present our own study, which is just a first (tiny) step in the direction of investigating the role of the language system in shaping meaning and cognition.

We should now have a first, although very general, picture of where our argument is going. In the remaining of this section we will walk step by step through this proposal, discussing its main strengths and possible pitfalls, some of our main points of interest and concerns.

[160] For a very interesting perspective on how we learn from language statistics cf., for example, Saffran (2002, 2003).

6.2.2. What distributional information cannot account for?

Let us begin by reviewing some of the main limitations of distributional representations of linguistic meaning if taken to account, by themselves, for meaning at a cognitive level of analysis. It seems important to point out, once more, that at this point of our discussion we are mostly concerned with the content of meaning and not as much with its format, nor with the organization of the semantic space, nor with the processes involved. At present, we are interested in discussing where our knowledge comes from and to what experiences do we owe such knowledge.

Distributional representations fall short on explaining the referential aspects of word meaning. We can read a lot of pages about pikas, and probabilistically infer many properties from knowing with which words the word pika is similar to (Erk, 2016), although without acquiring any (or very little) skills on how to recognize it if we ever actually were to encounter one. On this line of thought, it has been argued that, even if covariation among words is certainly related to meaning, meaning similarity, and psychological processing, distributional information do not have any causal role in meaning formation, but are rather a product of it (A. M. Glenberg & Mehta, 2008). This type of critique to distributional models was clearly expressed by Glenberg and Robertson (2000), but it can be much better understood if related to the much broader debate in the cognitive sciences about the *symbol grounding problem*: "How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?" (Harnad, 1990, p. 335). One of the main critiques to distributional representations is that the meaning of linguistic symbols (words) is given by a distribution over other linguistic symbols (contexts), in a never-ending regress that fails to establish a link with the referential content (but also with the perceptual and motor faculties) that allows us to interact with the world; internal manipulation of those symbols is never enough to establish meaning.[161]

The symbol grounding problem is directly connected to a second shortcoming of distributional representations, i.e. that of being essentially disembodied. In fact, the current debate on the cognitive role of distributional representations runs parallel to the broader debate between models of concepts that reduce their content to embodied sensory-motor information and models of concepts as symbolic and abstract entities (cf. Lenci, 2008).

[161] This could be seen as a variant of the famous "Chinese room" argument by Searle (1980). For a review of the argument as not necessarily cogent for symbolic approaches cf. Shapiro (2008).

Over the last decade, there has been an increasing body of work that explores whether sensory and motor information is a necessary part of semantic representation and processing, giving rise to a variety of approaches that fall under the more general notion of *embodied theories*. We have already briefly encountered embodied theories in chapter 3 (§5.3.1). Let us briefly recall some of the aspects with which our discussion is concerned.

According to early cognitive approaches, cognition was viewed as manipulation of symbols and the structure of the process and of the processor was more important than the kind of symbols it manipulated. The mental world builds a model of the external world by transducing perceptual inputs into other kinds of symbols. In line with symbol processing, early theories of semantic representation focused on how words are related to each other, i.e., to how the semantic system is organized. Conversely, under embodied approaches, there is only limited modelling of the external world and cognition is about real-world action rather than symbolic representation. Embodiment focuses on the content of cognitive representations and from that derives their organizational principles. Even if it remains clear that the environment has to be internalized somehow, embodied theories propose that, instead of transducing the signal into a symbolic format, the signal acquired through the senses is recreated when needed. In general, according to this view, concepts (and meanings as well) are not amodal, formal symbols, but rather inherently modal entities, represented (simulated and re-activated) in the same perceptual systems from which we acquire experience of their instances. The content of semantic representations is sensory and motor information (for a clear review and classification of embodied approaches cf. Meteyard, Cuadrado, Bahrami, & Vigliocco, 2012).

Over the years, increasing evidence has been collected suggesting that human concepts are strongly grounded and embodied in our senses and there is indeed empirical evidence that lack of sensorimotor grounding makes DSMs behave in ways that are (to some extent) different from those of humans (e.g.: A. M. Glenberg & Mehta, 2008; A. M. Glenberg & Robertson, 2000; Riordan & Jones, 2011). For example, DSMs often fail in the apparently simple task of labeling common concrete objects with their typical color (Bruni, Boleda, Baroni, & Tran, 2012). On the contrary, as we have discussed in the previous sections, it is similarly evident that DSMs can indeed predict a wide range of behavioral linguistic data and, hence, that purely intra-symbolic relations might indeed play a role in providing us with the relevant experience to build cognitive representations.

On the one side, it may sound rather intuitive that distributional representations easily account for one fundamental weakness of embodied theories, i.e. for how we can acquire and construct the meaning of abstract words or of words for concrete objects that we have never encountered and for which we are unlikely to have a specific sensory representation.[162] Some words do not seem to derive their meaning from grounding on embodied experience. Instead, their meaning seems to result from their relations to other words or concepts. Whereas the meanings of most concrete words are learnt in perceptual and manipulative contexts, it seems that the meaning of words like justice, truth, confidence, brotherhood, etc., are not necessarily derived from such concrete scenarios. Concrete concepts can also sometimes be learned from verbal definitions, or from implicit learning from context, rather than from grounding on experience.

On the other side, distributional representations are not symbolic in the exact same sense in which early cognitive accounts were. Distributional statistical (de Vega et al., 2008) symbols take as their units of analysis actual words in real context, rather than a language of thought as theorized by traditional symbolic approaches. Theoretically, the language of thought operates with mental symbols which, unlike words, are conceptual, abstract, and sometimes amodal, although they share with natural language some formal properties, such as the potential arbitrariness of symbols and syntactic rules. Furthermore, the statistical symbolism approach is more empirical than the traditional symbolic approaches. For instance, the impressive ability of DSMs to predict psychological phenomena emerges directly from the words themselves or, more accurately, from the statistical properties of words taken from natural language corpora and not from formal features of meaning postulated by researches. These substantial differences from traditional symbolic approaches position distributional semantics at the very margins of the symbolic versus embodied debate. In fact, statistical symbolists accept in principle that grounding on perception and action can happen, but they do not capitalize on the functional power of embodied processes in, at least, on-line comprehension of language. Instead, computations on symbols, words, or other

[162] We will not extensively discuss the shortcomings or the possible pitfalls of embodied theories of meaning, but we will rather mention just what may be relevant to our discussion. There are indeed multiple theories within the embodied literature that propose ways to overcome theoretical problems such as that of abstraction and of abstract words (for a very clear review of the debate between symbolic and embodied theories and some of the shortcomings of both, cf. de Vega, Graesser, & Glenberg, 2008). We are not arguing against the possibility that embodied approaches might eventually solve these as well as other theoretical riddles. Nonetheless, at least at the current state of research, it seems that an integration of the embodied and the distributional perspective (at least recognizing that both might play a role in shaping our conceptual knowledge) offers a solid and economic answer to many of these questions.

representations that are disembodied or not fully embodied can explain much of meaning in most language comprehension tasks.

Hence, in line with a now strong and affirmed paradigm of research, we argue that meaning (as we will shortly see, both in the language system and in cognition) is inherently multimodal, i.e. both embodied and symbolic. This statement needs to be accurately broken down and analyzed as it embeds two different, although related, claims.

6.2.3. Multimodality in language

The first side of the claim is that the meaning of words is both embodied and symbolic in language. On this line of thought, de Vega, Grasserer and Glenberg state that

> [l]inguistic symbols (words) are arbitrary in the sense that they do not resemble their referents except for the extreme case of onomatopoeia. Words are also abstract labels for categories or types. Finally, words are combined into high-order units by means of syntactic principles and patterns. All these are typical features of symbolic systems. Linguistic symbols prompt initial activations that are strongly embodied. The activation of linguistic symbols are not amodal but multimodal (2008, p. 435).

On the one side, words are arbitrary "abstract labels" with marked symbolic features; on the other side, when processed, they activate multimodal embodied representations. Statistical symbolists view words as pointers to modal representations grounded in embodied experience; in this sense, statistical computations are made on mental symbols and/or meaning is an abstract property emerging from words but not identifiable with words. This view resembles the already mentioned view according to which words can be regarded as cues to meaning: in a reformulated framework, although words may not be directly grounded in embodied experience, they activate embodied representations. Conversely, embodiment theorists emphasize that words are perceptual events with rich sensory features (visual, auditory, and motor) and that some of these features migrate to the encoded meaning representations. The choice between these two positions does not need to be either, or: both aspects contribute to how meaning is encoded in language.

According to the *Symbol Interdependency Hypothesis* (Louwerse, 2007, 2018; Louwerse & Jeuniaux, 2008), language encodes perceptual information and language users rely on language statistics in cognitive processes;

> language has evolved such that it maps onto the perceptual system, whereby language users rely on language statistics, which allow for bootstrapping meaning also when grounding is limited (Louwerse, 2018, p. 573).

One way in which we can learn and represent the world is by means of learning these intra-linguistic relationships. Instead of viewing language as either symbolic or grounded, language is organized so that it reflects embodied relations. Louwerse allows that many words, functioning as symbols standing for entities in the physical world, must be grounded through embodied semantic representations. However, words as symbolic representations also develop and maintain rich connections with each other. Thus, words may have grounded representations, but do not necessarily have to. Louwerse proposes that the network of symbolic connections supports efficient online language processing by allowing listeners to skip the enactment of embodied simulations. That is, if distributional structure is available, then language comprehension can rely solely on symbolic representations, without needing to invoke perceptual simulations unless deeper semantic processing is required. Given that language encodes embodied relations, in the comprehension process, comprehenders can benefit from interrelations between symbols because these are mapped onto relations in the real world.[163]

To this regard, our proposal is in line with the claim of the Symbol Interdependency Hypothesis, but we ought to explain the same outcome (words are both embodied and symbolic, and the language system provides efficient data to support language processing) along slightly different theoretical lines. We are not denying that much perceptual knowledge can actually be encoded in language, but we would argue that this fact can be explained within our theoretical framework in a way that slightly diverges from Louwerse's idea that language reflects embodied relations.

In relation to the world, the language system can be thought of as a web of pertinences that organizes our experience (cf. §§ 2.2.3-2.3.3 for a full discussion). This web of pertinences does not merely reflect the structure of the world, but at the same time carves nature at its joints. The way in which the system of language is organized is, at least to some extent, specific of language as a standardized web of inter-dependent intra-linguistic units. Moreover, we have been arguing that the level of the language system is never independent from the level of the actual individual

[163] For a review of recent works in philosophy, psychology, cognitive neuroscience, and computational modeling that endorse an interdependent view of intra-linguistic distributions and sensory-motor data cf. Andrews, Frank, and Vigliocco (2014).

language uses, and, hence, that it is always dependent on our individual psychological construction of the world, which depends upon the (also non-linguistic) way in which we perceive and categorize the world. Therefore, meaning in language is always constrained by the way in which we perceive the external (as well as our internal) world, and the content of our experience is always injected in the constant flow of language uses that make up the language system. Hence, the fact that language encodes perceptual features could be as well explained accordingly; the "language-specific" level of meaning is constrained but not fully determined by our non-linguistic organization of experience, and such constrains become especially visible in how words cue the reactivation of embodied representations. Language does not merely reflect our cognitive organization, but it partners up with in shaping our knowledge. In this way, the outcome is the same of that of the Symbol Interdependency Hypothesis, but the theoretical position underlying it slightly differs. Furthermore, our view allows us to theoretical join the fact that language encodes perceptual features (e.g. Louwerse & Jeuniaux, 2010) with the fact that, when DSMs are integrated with perceptual (mostly visual, but as well auditory) information, they actually perform better than only text-based models, suggesting that perceptual and textual information are tapping onto different aspects of meaning (Andrews et al., 2009; Baroni, 2016; Bruni et al., 2012; Bruni, Tran, & Baroni, 2014).[164] As we have argued multiple times already, distributional representations can be thought of as tapping onto the more "language-specific" relational level of meaning. This discussion allows us to point out a first way in which meaning is multimodal, embodied as well as symbolic.

6.2.4. Multimodality in cognition

The second part of the claim, clearly connected to the first one, is that meaning is multimodal at the cognitive level: the content of our representations is grounded in both action and perception as well as in our experience with language. There is a growing trend in cognitive

[164] A distinction can be made between DSMs models implemented with feature based representations, which are taken to proxy perceptual knowledge, (e.g. Vigliocco, Meteyard, Andrews, & Kousta, 2009) and (more recent) models implemented with visual or auditory features extracted directly from visual or auditory inputs (e.g.: Bruni, Boleda, Baroni, & Tran, 2012; Kiela, 2017; Kiela & Bottou, 2014). Although they use different techniques, both human feature-based and automatic-feature based multimodal DSMs (mDSMs) report increased performances. Therefore, in order to limit our discussion, we will treat them in general terms and avoid deepening the discussion into the various forms of mDSMs.

sciences to establish a common ground in which embodied cognition and distributional approaches to meaning could eventually meet. Barsalou et al. (2008), for instance, propose a dual model of meaning in which embodied simulation is accompanied by processing based on linguistic word co-occurrences, in the spirit of the distributional hypothesis, although linguistic distributions are assigned just a superficial role, with "real" deep semantic processing being carried out at the level of embodied representations. On this issue, Lenci (2008) concludes that "a promising line of research may come from assuming that between embodied cognition and the DH actually exists a sort of *division of semantic labour*, and that the empirical problem rather lies on how to divide their respective contributions in constructing meaning" (p. 25). Furthermore, based on the evidence that the information implicitly extracted from co-occurrence in text is qualitatively different from the information derived from sensory-motor experience, it has been argued that the integration of the two sources of information provides, at both theoretical and methodological level, the ground to explain a wide variety of language-related phenomena (both at an experimental as well as at a modeling level) that would be left out if the two sources of information would be kept separate.

Let us briefly review what distributional models of meaning can do if we feed them with perceptual and linguistic inputs at the same time, hence allowing us to talk about multimodal DSMs (mDSMs). How can we feed perceptual inputs to DSMs? Recent progress in computer vision has enabled us to extract vector-based representations of objects in natural images which are highly compatible with vector-based word representations. Vision is of course far from exhausting the ways in which language interacts through our senses with the outside world.[165] However, it is a central block of the perceptual experience of sighted people, and it can be taken as a reasonable starting point in the construction of multimodal computational models as well (for a comprehensive review and discussion cf. Baroni, 2016).

Cross-modal and multi-modal DSMs have proven to be able to deal with referentiality and grounding respectively. Cross-modal models attempt to translate linguistic representations into visual ones, and vice versa, in order to establish a link between words and the things they denote in the visual world. For example, given the distributional vector of a word and the picture of a certain scene, we would eventually be able, by converting the linguistic representation into a visual one, to tell whether the word represented by the distributional vector denotes an object that is

[165] For models that integrate text-based distributional semantics with auditory and olfactory information cf. Kiela and Clark (2015) and Kiela et al. (2015), respectively.

present in the depicted scene. For instance, Lazaridou and colleagues (2014) take up the challenging task of associating a word with an object when both are novel for the system, a realistic scenario in human word learning. When a human reads the sentence "There is a cute hairy wampimuk sitting by the tree", he or she will have a good idea of what a "wampimuk" will look like even if he or she has never seen it (or indeed, if it doesn't exist at all). The model presented by the author was actually able to learn the mapping from text to vision. Cross-modal mappings are usually evaluated through leave-one-out experiments, where the objective is to map from one space to the other and retrieve the correct image or word, without ever having seen it before. Cross-modal semantics is a recent research arena where many questions are still open and unexplored (for a review and discussion cf. Kiela, 2017).

As for multimodal DSMs, researches have been combining linguistic and visual representations in order to obtain well-rounded representations of semantic properties. That visual information could enhance textual representations was already pointed out by Landauer and Dumais (1997) at the very beginning of the distributional enterprise, but practical implementations were described only later (e.g.: Bruni, Tran, & Baroni, 2011; Feng & Lapata, 2010). The general mechanism is that, given the textual and visual vectors for the words/concepts of interest, a multimodal fusion function returns a new set of vectors combining textual and visual information, and thus residing in a multimodal space. A very simple but effective concrete example of fusion is to just concatenate the textual and visual vectors associated to a word, deriving new vectors that have both textual and visual dimensions (Bruni et al., 2011; Kiela & Bottou, 2014).

In conclusion, multimodal and cross-modal DSMs tackle onto those aspects that have been often pinpointed as pitfalls of the distributional approach and allow to draw an interesting conclusion: multimodal models, hence multimodal representations, perform better that single word or single visual information models on a wide variety of language related psychological tasks. Multimodal semantic models correlate with human knowledge better than the approximations given by linguistic patterns or by perceptual information alone. This line of work, still in its early stages, is highly promising for providing further insights into what kinds of knowledge perception and language offer to the learner. It bears mention that although the performance of these models shows that it is (in principle) possible to learn these semantic embeddings from the input and to integrate information from different sources into one single representation, it does not necessarily follow that people's semantic knowledge is learned or structured in the same fashion. The main

conclusion of this portion of our discussion is that there seems to be an increasing support for the idea that the way in which we construct meaning in the mind is by integrating multiple sources of information into a multimodal representation of our experience. It is nonetheless clear that we are still a long way from identifying the precise mechanisms that underlie acquisition from experience and even a longer way from being able to precisely define the contribution of each element to the formation of our knowledge.

However, interesting insights that point to the same conclusion come from the psycholinguistic literature. Hence, we are now moving from the implementation of computational models to the analysis of psycholinguistic behavior data that could help us shed light on the content of the information that we, as humans, are able to extract from language. If we assume that cognitive representations are multimodal, and that language is one source of information that feeds our process of meaning formation in the mind by being an external input as much as an internal construct, one straightforward question would be: *what kind of information does the language input (the distributional properties of words) actually provide us with*?

There is now an increasing body of evidence from psycholinguistic studies that distributional information actually plays a role in the way in which we learn the meaning of novel words. It is well established that the context in which an unfamiliar word occurs is an important determinant in learning its meaning, and it is sometimes also evident that context provides the sole means for establishing it. Contextual cues also play an important role in consolidating the meaning of newly-learned words. The more exemplars of a word in its context of use are encountered, the more its meaning can be refined and delimited, especially if one has some prior knowledge of the discourse or passage topic.[166] For instance, McDonald and Ramscar (2001) manipulated subjects' contextual experience with marginally familiar and nonce words. Results showed that similarity judgements involving these words were affected by the distributional properties of the contexts in which they were read. On quite the same line of research, Lazaridou et al. (2017) show that minimal distributional evidence from very short passages suffices to trigger successful word learning in subjects, testing their linguistic and visual intuitions (hence, intuitions in different modalities) about the meaning of a novel word. Studies in developmental psychology have been endorsing a positive role of the linguistic context in which words are embedded in the implicit acquisition of the meaning of novel words for quite a while now, providing convincing evidence

[166] For a discussion of the role of contextual diversity in word learning cf. e.g.: Johns, Dye and Jones (2016)

that language experience indeed plays a role in language development (Nagy, Herman, & Anderson, 1985; Swanborn & De Glopper, 1999; Hulstijn, J.H., 1992). Research in language development, even outside of the distributional framework, payed substantial attention to the role of the various factors that have shown to influence learning word meaning from context, such as frequency of the words in texts (J. R. Jenkins, Stein, & Wysocki, 1984), target-context ratio (Browne, 1989; Redington, Chater, & Finch, 1998), proximity of the cue to the unknown word (Nagy et al., 1985), morphological transparency of target words (Swanborn & De Glopper, 1999), and salience (C. M. Brown, 1993). Thus, so far, we have been able to establish that both adults and children are good distributional learners. However, acknowledging this evidence does not take us any closer to determine what kind of information we get from language, besides the general intake that we are able to learn the meaning of novel words and yet with very little agreement on what aspect of language may be driving this acquisition.

To this regard, some further suggestions come from the debate in the cognitive sciences. Let us briefly review two of the most salient proposals about what the inputs from language might allow us to achieve: linguistic inputs seem to be particularly suited for the acquisition of abstract words and for promoting abstraction and augmenting perceptual representations to make them more categorical.

Vigliocco and colleagues (2009) argue for the joint contribution of (perceptual and emotional) experience and linguistic experience in shaping meaning. More to the point discussed here, the authors suggest that differences between concrete and abstract word meanings (as well as within concrete and within abstract word meanings) arise as a result of the proportion and exact type of experiential and linguistic information from which they are derived. The apparent dichotomy between concrete and abstract[167] word meanings arises in the statistical preponderance of sensory-motor information underlying concrete word meanings and the preponderance of affective and linguistic information underlying abstract word meanings. Similarly, the linguistic nature of semantic representations for abstract words is argued in models such as the dual coding theory of Paivio (1986) according to which, abstract words would be represented solely in terms of a "verbal code". Paivio's dual code theory argues that concrete words are processed both by the

[167] Concreteness is defined by the authors as "an operationalization of the distinction between entities and events that exist in the physical world and entities and events that exist in the human mind", in contrast to imageability, which is "an operationalization of the relevance of especially visual sensory properties of entities and events" (Vigliocco et al., 2009, p. 232).

symbolic (verbal) and the embodiment (imagery) systems, whereas abstract words are processed just by the symbolic system. In conclusion, the general takeaway is that linguistic information seems to be implicated especially in the acquisition and processing of abstract meaning.

However, as we have seen throughout our discussion, if linguistic information seems especially suited for abstract words learning, linguistic input is surprisingly rich and informative even at the level of our knowledge about concrete entities, both because language encodes perceptual information and because the linguistic categorization of experience is not completely derived by our sensory experience with it, but rather provides us with an different layer. Hence, a further question follows: is there a systematic difference in the kinds of semantic knowledge that may be formed from language versus from perceptual experience? Does language play a specific role in the way in which we acquire and conceptualize meaning? As we have seen, recent advances within the modeling enterprise suggest that the experience provided by language is qualitatively different from the experience provided by perception and action. On this line on investigation, some more insight comes from studies in which linguistic factors are experimentally manipulated while people attempt to learn new categories or use existing knowledge to recognize or make inferences about familiar categories. Under the influence of language, semantic knowledge may become more categorical with consequences for behavior ranging from basic perception to reasoning (Lupyan, 2012a, 2016; Lupyan & Bergen, 2016; Perry & Lupyan, 2013). Language seems to promote abstraction: we perceive specific objects and events, but we talk about them categorically. As a consequence, learning and using verbal labels appears to augment perceptual representations to make them more categorical. For instance, studies of category learning in adults have shown language to facilitate learning of new categories (Lupyan & Casasanto, 2015) and that language continues to aid categorization even of previously learned very familiar items (Lupyan & Thompson-Schill, 2012). In a nutshell, adding linguistic experience can enhance categorization; for instance, results from Edmiston and Lupyan (2015) suggest that language activates visual representations that are partly constitutive of visual knowledge, but in so doing, augments them into a more categorical form than when ostensibly the same representations are activated by nonlinguistic inputs. Whether these suggestions about the role of language in acquiring abstract content or in aiding abstraction and categorization would be proved correct or not, we are still left with our original question about what in the language inputs provide us with the relevant information for giving rise to our multimodal representation of the world.

6.2.5. The hypothesis: the structure of language as a main source of meaning construction

At the current stage of the research on this topic, this question remains mostly unanswered. At least in light of our review of the literature, this is not due to a lack of interest for the role of language in cognition, but rather to the inherent complexity of dealing with meaning in a unitary framework. A considerable portion of approaches to semantics still tends to isolate specific aspects of meaning and to build upon those. In this last chapter, we have mostly reviewed proposals that are moving in the direction of providing an integrated view of meaning that could encompass and intertwine intuitions from linguistics, data and methods from psychology, and proposals and implementations from modeling approaches. A clear benefit of these integrated views is that they have the potential to overcome the discrete disciplinary boundaries that have often been the hallmark of the research on meaning. Our proposal, in line with these attempts, tries to build a profitable line of dialogue at the intersection of linguistics and cognitive approaches to meaning.

Moreover, even when embracing a multidisciplinary approach and a unitary view of meaning, the difficulty in answering the question about what in language may actually be the source (or sources) of meaning construction is due in part to the empirical difficulty in assessing the provenience of our knowledge as, as we have argued, representations in both language and cognition are inherently multimodal. However, as we have already pointed out, there seems to be aspects of the way in which we organize our knowledge that lean more to the side of language or more to the side of cognition. There seems to be aspects of cognition that shed light onto the content and organization of meaning in language (embodied theories give us a precisely to the point account of this) and, conversely, aspects of the organization of meaning at the "language-specific" relational level provide us with the theoretical insights for understanding some fundamental aspects of our cognitive organization. Hence, although there seems to be a specialization of the two in providing us with the relevant information in our process of content formation, the line dividing the two (or even just a possible imaginary line on which the two converge) is, at best, hard to define. But is this research program actually deemed to fail? As our review of the literature demonstrates, enlightening attempts in this direction are already in place and are progressively converging towards similar intuitions and gaining consensus across a wide variety of disciplines and approaches.

Our approach sets itself in this space of convergence and aims at outlining a possible way of deepening our understanding of the role of language – of the "language-specific" level of meaning - in shaping cognition. The final step of our proposal intends to spell out a possible line of inquiry into the study of language as a source of information in cognition. We want to draw attention to one aspect of language that has often been underestimated in the cognitive study of meaning, but that we have tried to re-establish as a fundamental facet throughout our discussion. Meaning in language is organized by means of a web of relationships that gives rise to the language system and to the "language-specific" aspects of linguistic meaning. We have already insisted on this point multiple times. How is this relevant to our discussion here? Our suggestion is that a possible line of inquiry would be to experimentally evaluate the possibility that one central aspect of what we learn from language is, together with specific contents, the relational system of meaning. According to our perspective, content and structure are inherently related and inseparable, in the sense that meaning at a linguistic level, although limited by our biological nature, is always a matter of the relationships it establishes with other units in the system. Our proposal, which is backed up by the evidence reported so far and allows us to make interesting and testable hypothesis and predictions about the role of language in cognition, is that what language brings to the table is its systemic organization, its web or relationships and of pertinences that do not completely overlap with what we get from other aspects of our experience. Distributional approaches can be thought of as providing the way into the operationalization of the systemic aspects of meaning in language, making it suitable for experimental manipulation. At its core, our idea is that testing the role of the distributional information in learning meaning approximates the study of the role of the language system in shaping our cognitive representations.

Indeed, this is easier said than done, and a word of caution is definitely in order. In the first place, due to the multimodality of representations in language (i.e. the fact that words encode perceptual features and that, at the same time, they can essentially be viewed as cues to much broader frames of encyclopedic knowledge), it would be misleading to claim that distributional representations are the straightforward representation of the system of language. We envisage future research in both psycholinguistics and computational linguistics to further shed light onto the nature and properties of distributional representations. As our review illustrates, at the current stage of research, there is solid evidence suggesting that distributional representations can be interpreted as (at least) tapping onto the relational aspects of meaning; therefore, we will assume

that distributional information can at least proxy the systematicity of linguistic meaning. In the second place, at the level of linguistic analysis, distributional semantics implies that there is a continuum between semantic, syntactic, and pragmatic aspects of meaning. Hence, manipulating the linguistic environment in experimental conditions often implies simultaneously manipulating all three aspects without a clear way to identify the role of each. In the third place, we are indeed very far from a full-fledge experimental research paradigm that could shed light onto the role of the language system as a building block for meaning. We envisage future research to overcome the limitations of our current approach, by establishing a way to manipulate the linguistic environment so to isolate, for example, the role of language-based pertinences is shaping our categorical representations of the world.

On this line of inquiry, we propose an exploratory study that aims at evaluating the limits of our ability to learn semantic content from distributional information.

6.3. Rapid learning of word meaning from morpho-syntactic and distributional cues

This paragraph reports on the results of a project carried out in collaboration with Gary Lupyan (Department of Psychology, University of Wisconsin-Madison), started during a vising scholar research period from January to June 2018 and still on-going. The results were previously presented at the 41st Annual Meeting of the Cognitive Science Society (24-27 July 2019) and the paper was published in the conference proceeding.

When we look up the meaning of an unknown word in the dictionary or when someone explicitly defines the meaning of a word to us, it is rather obvious that language is implicated in the way in which we acquire and construct the meaning of a certain word. At the same time, when taking the TOEFL Test of English as a Foreign Language the ability to infer meaning from the linguistic context is essentially considered as one of the skills required for a second language learner of English to score highly in the reading section. Players of the online interactive game *The Gostak* are able to navigate the game's alien world by making sense of a text adventure written almost entirely in a gibberish language, the Gostakian, essentially English except most of the nouns, verbs, adjectives, and adverbs are completely new and might not have an exact equivalent in English. Thus, the linguistic context in which the unknown words appear seems to be an information source from which the meaning of new words can be explicitly or implicitly learnt. However, the mechanisms by which we are able to infer linguistic meaning from its linguistic context are still an open and debated issue. What aspects of language allow us to acquire the meaning of words from the linguistic context? In this work we focus on the role played by morpho-syntactic information provided by the linguistic context in constructing linguistic meaning. Traditional accounts have mostly been interested in assessing the role of semantic informative contexts in the acquisition of novel meaning in children or in L2 learners. In developmental psychology, it is now widely recognized that learning words from context is a major way of vocabulary acquisition and that this often occurs incidentally; this is, school children show reliable effects of learning word meaning from exposure to written texts without direct instructions (Kuhn & Stahl, 1998; Nagy, Anderson, & Herman, 1897; Nagy et al., 1985). Over the last few decades, studies have focused on the relevant methods in accessing partial acquisition of linguistic knowledge (Hulstijn, J.H., 1992; J. R. Jenkins et al., 1984; Swanborn & De Glopper, 1999) and have established the incremental nature of word acquisition (Fukkink, 2005; Fukkink, Blok, & de Glopper, 2001). As we have already mentioned, research has focused on the factors that have been

shown to influence learning from context; frequency of the word in the text, ratio between the whole text and the target words, word salience, informativeness and variability of the context are all aspects shown to affect the probability of learning an unknown word while reading. As for the role of syntactic structures in shaping the meaning of novel words, developmental studies have largely addressed the role of syntactic-bootstrapping: pre-school children have shown to use both the morphology and the syntax (in particular noun phrases and verb phrases) in which a word is placed to narrow down or constrain the meaning of a certain word (R. W. Brown, 1957; Naigles, 1990; Naigles & Swensen, 2007). Syntactic bootstrapping is a probabilistic rather than categorical procedure based on the correspondence between syntactic and semantic meaning and operates in concert with the observed visual scene and with pragmatic information. Although the majority of the studies have been concerned with the investigation of specific syntactic nominal and verbal structures (e.g. transitive/intransitive) and related interpretative frames (e.g. agent/patient), a few studies have addressed the involvement of language-specific morphological patterns in learning the meaning of new words proving that school-age students use morphological analysis to determine the meaning of unknown words (McCutchen & Logan, 2011; Wysocki & Jenkins, 1987). In addition to the meaning provided by specific syntactic frames, the distributional properties of language itself are informative about the structure and meaning of language. The exploitation of regularities in the linguistic environment has been put forward to explain how language learners accomplish various tasks, from segmenting speech stream into words (Saffran, Aslin, et al., 1996) to actually acquiring syntactic knowledge (Redington et al., 1998).

Morpho-syntactic information and distributional learning: the Jabberwocky word learning paradigm

When George in *A clockwork orange* says “Dim was a lot more starry and grey and had a few zoobies missing as you could see when he let out a smeck, viddying me, and then my droog" we may not understand the possible English equivalent of each word, but we most probably know what the overall meaning of the sentence is and the feeling that it evokes. As we interpret this sentence, however, we are generally helped by either the visual and auditory stimuli provided by the movie or by a larger semantic coherent context provided by the book. But can we still infer the meaning of the unknown words if our only cues are morphological and syntactic ones in the

linguistic context itself? In our studies, we draw inspiration by the above-mentioned approaches and assumptions to target the role of morpho-syntactic information in the acquisition of semantic meaning. Working with English native speakers, we present participants with a moderately complex text in which we replaced all open-class words with nonce words that respect the English phonology and we maintained unchanged only syntactic structure, morphological variations and stop-words of the English language. We test participant on inferences about the part of speech of the target words as well as on more specific meanings. The use of gibberish language or of very rare English words in studies focused on distributional semantics and learning from context is not new: McDonald and Ramscar (2001) test the distributional hypothesis by manipulating the linguistic context surrounding very rare or nonce words showing that similarity judgments involving the target words are affected by the distributional properties of the manipulated linguistic environments. Ouyang, Boroditsky and Frank (2017) argue for semantic coherence and availability of semantic meaning as important factors in distributional learning of nonsense words and Lazaridou, Marelli and Baroni (2017) investigate the generation of novel concepts in adults driven by the acquisition of a novel nonce words from the linguistic context. In parallel with this line of research, we use the gibberish paradigm to test the effect of syntactic and morphological organization of language in shaping the meaning of words. If morphology and syntax ought to show a significant impact on inferring semantic meaning from linguistic context, it would in the first place show further support for the learning from context claim as well as for the distributional hypothesis. As it will become more evident in a detailed outline of the set-up of the studies, the majority of the information available to participants is of a distributional kind. In the second place, providing evidence for the role of morpho-syntactic cues in word learning endorses the idea that syntactic structures and morphological aspects of the sentence provide speakers with important guidance in constructing meaning and that themselves are bearers of semantic meaning. Furthermore, by testing for both part of speech and inferences about the specific meaning of the target words our studies might shed light on some of the mechanisms involved in the acquisition of language from the linguistic environment as well as on the kind of meaning supplied by morpho-syntactic cues. Our main goal is to investigate the limits of people's ability to learn word meanings from linguistic context. We do this by exposing adult English speakers to contexts varying in informativeness ranging from fully informative contexts (passages of real English text containing real words unknown to most of our subjects), to highly sparse contexts that contain English

morpho-syntactic cues (e.g., verb endings), but in which all content words have been replaced by nonsense words. We test people's knowledge of both part of speech information (the sole focus of much of the classic work on syntactic bootstrapping), and knowledge of more specific aspects of meaning.

Experiment 1

In experiment 1, we tested the role of the linguistic context in inferring word meanings for real, but rare English words (e.g., "kine") that were presented in informative contexts and for nonce words (e.g., "stronk") placed in highly sparse contexts which were stripped of almost all meaningful words.

Participants

We recruited 114 participants from Amazon Mechanical Turk (52 Males of average age = 37; 62 Females of average age = 39). 68 of these participated in the main word-learning experiment (32 in the nonsense-story condition and 36 in the real-passages condition) and 46 participated in the salience-norming task.

Materials

Participant were randomly assigned to one of two conditions: a nonsense-story condition and a real-passages condition. For the nonsense-story condition, all participants were exposed to the children's story Why the cricket chirps2. Participants were not provided with the story's title or any information about its content. Of the 604 total to-kens in the story, 169 were open-class words. We replaced all the content words with nonce words taken from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). These words were created from orthotactically legal bigrams, onsets, and bodies. Fig. 1 shows a part of the resulting story. Of the 136 word types in the story that were replaced with nonce words, we selected 9 to serve as targets for later testing. Participants did not know ahead of time which words would be tested. The target words varied in frequency, occur-ring between 2 and 18 times, and parts of speech: 4 nouns, 3 verbs, 2 adjectives. Derivational and inflectional morphemes in the story were limited to a small number of cases (see Table 1 for morphological variation present for each target word). For the real-passages condition,

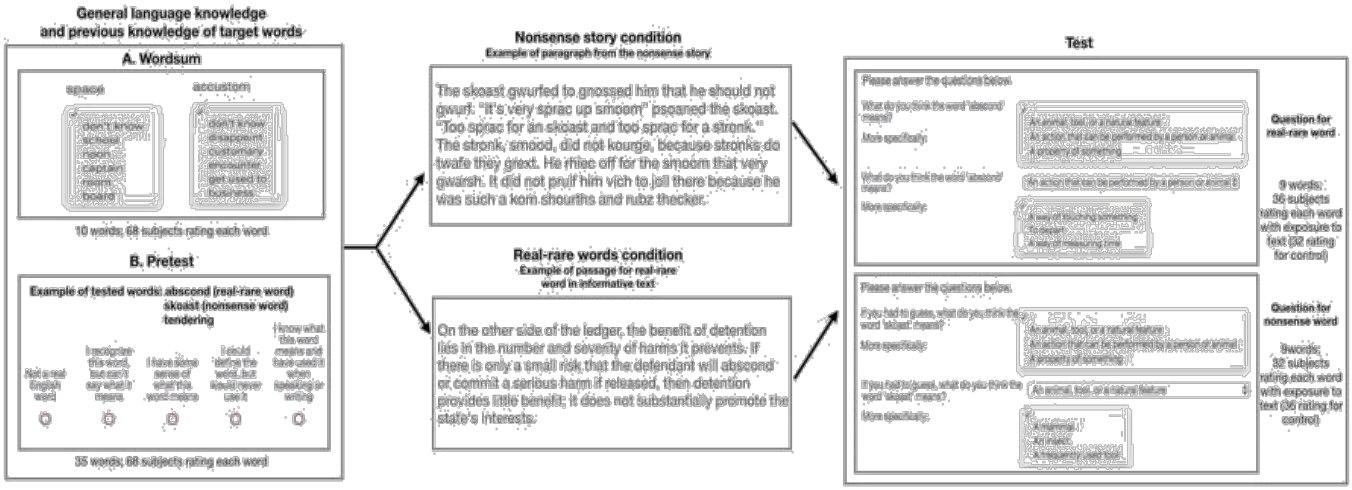


Figure 22. Schematic for experiment 1

we matched each of the 9 target words in the nonsense-story condition with real, but rare English words that were unlikely to be familiar to our participants (e.g. "ratoon", "pronk, "rawky"; henceforth real-rare words). For each word, we selected 3-4 sentences in which the word appeared from the Corpus of Contemporary American English (COCA) and other online sources to serve as the context (see Fig. 22). Participants' word knowledge in all conditions was tested using "drill-down questions" de-signed to be sensitive to partial word knowledge. The first question provided three options for part of speech and the second question had participants choose between three word meanings all within the chosen part of speech: 1 fully correct, 1 partially correct, and 1 incorrect (see Fig. 22).

Quantifying word salience

We expected that people's ability to infer word meanings would be influenced by the frequency with which the target word occurred in the passage. But we also suspected that aside from frequency, performance would also be related to a word's salience (C. M. Brown, 1993). There is no single definition of salience, but intuitively, a word is salient to the extent that it communicates the central point of a story. For example, words naming the actions performed by a central character are more salient than words describing aspects of the environment that a non-central

Target word	Frequency	Salience	Derivational Morphology	Inflectional Morphology
fly	18	6.59	1 derivational form (-er)	2 inflected forms (-ed; -ing)
cricket	16	31.69	none	1 inflected from (-s)
cold	7	5.69	none	none
wing	6	8.26	none	1 inflected from (-s)
fast	6	3.76	none	2 inflected forms (-er; -est)
chirp	5	8.49	none	3 inflected forms (-s; -ed; -ing)
ant	5	6.91	none	none
hop	5	1.06	none	none
snow	5	2.56	none	none
rub	4	1.90	none	2 inflected forms (-ed; -ing)
listen	4	1.47	none	none
warm	4	1.31	none	none
tree	3	0.44	none	1 inflected from (-s)
frozen	3	4.82	none	none
ground	2	0.00	none	none
owl	2	10.35	none	none

Table 2. Frequency, salience, and morphological variation of target words

character inhabits. We quantified the salience of each target word as the likelihood that it would be recalled by people who read the original (unaltered) story. We recruited 46 participants from MTurk to read the original story and then had 1 minute to recall all the words they could remember occurring in the story. Salience for each word was defined as the sum of the weights that the word obtained each time it was listed: the weight was calculated as exponentially decreasing in accordance with the order in which words were listed by participants [(for each time the word was listed) weight = (0.75ˆ(word n -1))] (see Table 2 for frequency and salience of each target word).

Procedure

General procedure is shown in Figure 1. At the start of the task, participants completed a 10 item vocabulary test (Wordsum; Malhotra, Krosnick, & Haertel, 2007) and a pretest gauging familiarity with the target words. Participants were then randomly assigned to the real-passages or the nonsense-story condition. Those assigned to the real-passages condition saw each (meaningful) context and answered the two vocabulary questions for each of the 9 real-rare words (in random order). Each word was tested immediately after being presented in its context. The group was then tested on nonce words (skoast, etc.). In contrast, participants in the nonsense-story condition saw the entire 604-word nonsense-story and were then tested on the real-rare words, and then on 9 of the target nonce words (in random order). This design allowed each condition to serve as the control for the other condition. Subjects in the nonsense-story condition were asked to infer the meaning of the real-rare words without exposure to the passages and vice versa. As an attention check, scattered among nonce and real-rare words were questions about the meaning of familiar words (e.g. “little”, “green”).

Results and Discussion

We analyzed the data using logistic mixed effects models. In the initial analyses, we treated partially and completely correct scores for specific meaning as the same (i.e., a binary contrast between an accuracy of 0 and 0.5/1). Figure 23 shows a clear interaction between condition (real-passages, nonsense story) and word-type (real-rare, nonce). This interaction was present both for part of speech [$z = 7.3$, $SE = 0.28$, $p < .001$] and word meanings [$z = 9.1$, $SE = 0.29$, $p < .001$] [accuracy ~ type of word*condition+(1jsubj id)]. Participants in the real-passages condition were significantly more accurate at selecting meanings corresponding to the correct part-of-speech of real words compared to participants in the nonsense-story condition (i.e., those not exposed to the real passages) [$z = -3.7$, $SE = 0.39$, $p < .001$]. They were also better at choosing the more correct specific meanings [$z = 4.6$, $SE = 0.4$, $p < .001$] [accuracy (for specific meaning or for part of speech) ~ condition+(1jsubj id)+(1jword)]. Note that above chance performance for the real-rare words is expected even without being exposed to real passages because some participants already know the meaning of these words. Not surprisingly, accuracy on real-rare words was positively associated with greater vocabulary as assessed by Wordsum. This was true both for part of speech [$z = 2.8$, $SE = 0.17$, $p = .005$] and specific meaning measures [$z = 3.7$, $SE = 0.19$, $p < .001$] [accuracy (for specific meaning or for part of speech) ~ condition*wordsum score+(1jsubj id)+(1jword)]. Greater familiarity with the target real-rare words (pretest) was posi-tively related to selecting the correct part of speech [$z = 2.1$, $SE = 0.16$, $p = .033$] and specific meaning [$z = 2.9$, $SE = 0.16$, $p = .004$]. The results for the real-rare words tell us what we already knew – people can infer word meanings from seeing the words in context. We now turn to the nonsense-story condition. Recall that these nonce words were seen in the context of a 600+ word story in which all content words were replaced by nonce words. Participants exposed to this extremely sparse context had significantly higher performance in inferring the correct part of speech for the nonce words [$z = 3.9$, $SE = 0.23$, $p < .001$] and in choosing the more correct meanings [$z = 4.4$, $SE = 0.25$, $p < .001$]. The benefit from reading the nonsense story was not limited to just helping people choose the correct part of speech. Restricting the analysis to only the trials on which participants

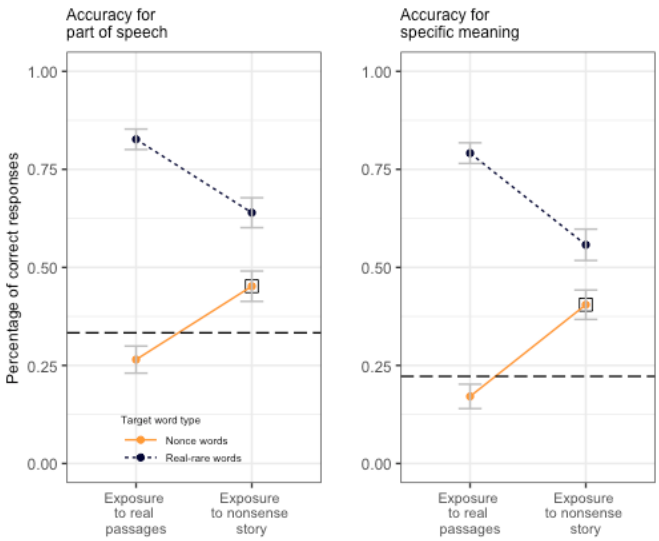


Figure 23. Accuracy for type of target word for Experiment 1. Horizontal dashed lines indicate chance-level. Error bars indicate +/-1SE of the mean. Significant effect of context exposure for nonce words in the nonsense-story is marked by a squared shape around.

chose the correct part of speech, we find that exposure to the nonsense-story still led to higher accuracy [z = 2.9, SE = 0.35, p = .004] [accuracy_part of speech ~ condition+(1jsubj id)+(1jword)]. Neither Wordsum nor Pretest scores predicted performance for nonce words. Word frequency and word salience likewise did not predict performance (z<1, but see Exps. 2 and 3).

Experiment 2

Experiment 1 showed that people can benefit from very sparse contexts. One shortcoming of the study was that participants in the real-passages and nonsense-story conditions were exposed to contexts in a different way (short passages vs. one long story; tested after each word vs. tested on all words at once). In Experiment 2, we test participants in the same way on all the same words, varying just the informativeness of the context. We were also curious about whether it mattered how word knowledge was assessed. Instead of using the "drill- down" format that first asked about part of speech, we used a more standard multiple-choice text, presenting all the 9 options for each word at the same time. Lastly, we parametrically varied the informativeness of the context by replacing different proportions of content words with nonce words.

Experiment	Condition Name	Type of Text	Type of Target Word	Example Words	Question Test Type
Experiment 1	Real-passages	Real-rare target words; informative text	real-rare words	ratoon; pronk; rawky	drill-down
Experiment 2	Nonsense-story	Nonce-target words; nonsense story	nonce words	stronk; sprac; crex	drill-down
	Real-text	Real-rare target words: real story	real-rare words	auscult; lollop; smuir	multiple choice
	Real-words-nonce-context: 40%	Real-rare target words; 40% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	multiple choice
	Real-words-nonce-context: 60%	Real-rare target words; 60% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	multiple choice
	Real-words-nonce-context: 90%	Real-rare target words; 90% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	multiple choice
	Real-words-nonce-context: 100%	Real-rare target words; 100% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	multiple choice
	Control group	No exposure to story	real-rare words	auscult; lollop; smuir	multiple choice
Experiment 3	Real-words-nonce-context: 100%	Real-rare target words; 100% of context replaced with nonce words	real-rare words	auscult; lollop; smuir	drill-down
	Control group	Nonsense story with all words replaced with nonce words	real-rare words	auscult; lollop; smuir	drill-down

Table 3. Summary of type of context, type of target word and methods is assessing word meaning in each experiment.

Methods

Participants were randomly assigned to one of six conditions (see Table 3 for a summary). In the real-text condition we exposed participant to the original Cricket story, in which we replaced only the target words with real-rare English words. Thus, the linguistic context was still informative (i.e., the target words were surrounded by meaningful words) but it did not directly aim at communicating the meaning of the target words, e.g., *'You should find some shelter from the cold night,' said the smew. The mitius did not auscult, because mitiuses do whatever they want. He decided to rest on a pile of twigs.* We progressively decreased the information pro- vided by the context by replacing various proportions of the remaining content words in the story with nonce words (40%, 60%, 90%, 100%). Lastly, we included a control group that was tested on their knowledge of the real-rare words without seeing any prior context.

Participants

We recruited 246 participants from Amazon Mechanical Turk (112 Male of average age = 35; 132 Females of average age = 38). 38 participants were assigned to the real-text condition, 36 to the 40% condition, 39 to the 60% condition, 42 to the 90% condition, 52 to the 100% condition and 39 to the control condition.

Procedure

Participants in each condition were initially tested on Wordsum and Pretest (following the procedure for Experiment 1) and then randomly assigned to one of the six conditions described above. All participants were then presented with the same 12 multiple choice questions (9 options per question) to assess their knowledge of the real-rare words.

Results and Discussion

Results are shown in Fig. 17. Participants exposed to the full story clearly benefited in inferring both part of speech [$z = 5.9$, $SE = 0.22$, $p < .001$] and specific meaning [$z = 5.6$, $SE = 0.25$, $p < .001$] of real-rare words [accuracy~overall ~ control vs full story +(1|subj id)+(1|word)]. Similar results were found for the 40% and the 60% conditions, in which participants showed compelling effects for both part of speech [$z = 4$, $SE = 0.22$, $p < .001$] and specific meaning [$z = 3.8$, $SE = 0.25$, $p < .001$]. In contrast, when 90% or 100% of content words were re- placed with nonce words, no significant benefit was observed for either part of speech [$z = 1.2$, $SE = 0.2$, $p = .226$] nor specific meaning [0.77, $SE = 0.23$, $p = .444$] results (Fig. 24). Frequency of occurrence in the story was positively associated with accuracy for the real-text condition, [$z = 2.7$, $SE = 0.16$, $p = .008$] [accuracy ~ condition*frequency+(1|subj id)+(1|word)] and salience [$z = 2.3$, $SE = 0.16$, $p = .023$] [accuracy ~ condition*salience+(1|subj id)+(1|word)] in predicting accuracy for specific meaning. More frequent and more salient words benefited more from context. Similar effects were found in the 40% and 60% conditions for accuracy on specific meaning [frequency: $z = 3.8$, $SE = 0.16$, $p < .001$; salience: $z = 2.3$, $SE = 0.16$, $p = .023$] [accuracy ~ condition*wordsum scores+(1|subj id)+(1|word)]. Controlling for pretest scores, greater vocabulary knowledge (Wordsum) was positively associated with accuracy for both part of speech [$z = 2.7$, $SE = 0.12$, $p = .006$] and specific meaning [$z = 2.7$, $SE = 0.14$, $p = .007$]. Similarly, previous word knowledge was positively associated with part of speech accuracy [$z = 2.4$, $SE = 0.084$, $p = .018$] and specific meaning accuracy [$z = 2.9$, $SE = 0.086$, $p = .003$] [accuracy ~ condition*pretest scores+(1|subj id)+(1|word)]. These associations parallel the findings of the real rare words condition of Experiment 1. Exposure to a story in which 40%-60% of content words were replaced with nonce words still allowed participants to learn something about meanings of words occurring in the story. There were two noteworthy differences between the results of Experiment 1 and 2. First, unlike

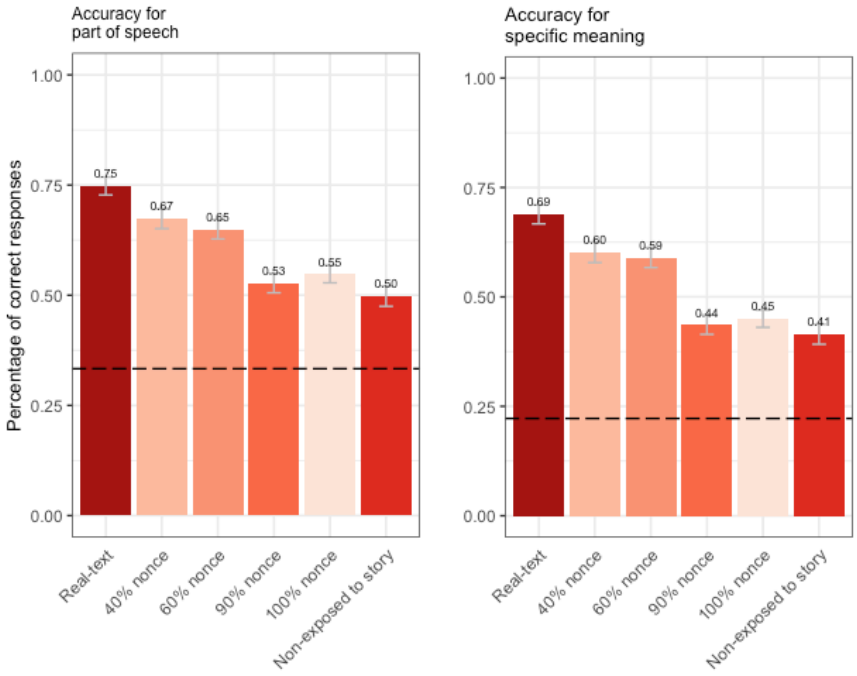


Figure 24. Group performance for Experiment 2. Horizontal dashed lines indicate chance-level. Error bars indicate +/- 1SE of the mean.

Experiment 1, participants' ability to benefit from the story context was positively associated with the frequency with which the word occurred in the story and the word's salience. These relationships may stem from people's greater baseline knowledge of the (real-rare) target words. Second, exposure to a story in which all the content words with nonce words did not lead to greater-than-baseline performance on the word test. Experiment 3 was designed to better understand this difference.

Experiment 3

In Experiment 1, we found a significant effect of context exposure in inferring the meaning of nonce words from nonsense context. In Experiment 2 we assessed knowledge of real-rare words instead of nonce words, and embedded them in contexts of varying informativeness. Contexts in which 40-60% of content words replaced with nonce words were helpful, but those in which more (90%-100%) of content words were thus replaced, were not helpful. How do we reconcile this difference? Aside from testing nonce words vs. real-rare words, Experiments 1 and 2 differed in the way word knowledge was assessed. Experiment 1 first asked about part-of-speech. Experiment

2 presented all the meaning choices together, intermixing meanings from different parts of speech. We reasoned that explicitly asking people about parts of speech (which are more directly bootstrapped by morpho-syntactic cues) may make it easier for people to subsequently access more specific aspects of the word's meaning. In Experiment 3 we tested the effect of exposing people to a nonsense-story containing real-rare words as the real-words- nonce-context (100%) condition of Experiment 2, but using the drill-down question format of Experiment 1.

Methods

Participants were randomly assigned to either the real-words-nonce-context: 100% or to a control condition. In the real-words-nonce-context, participants were exposed to a nonsense-story containing the real-rare target words and tested on those real-rare words (as in the 100% condition of Experiment 2). Participants in the control condition were shown a story with all nonce words (as in the nonsense- story condition of Experiment 1) but at test were asked about the meaning of the real-rare target words of the real-words- nonce-context condition (i.e., they were asked about words they did not see in the story).

Participants

We recruited 81 participants from Amazon Mechanical Turk (39 Male of average age = 37; 41 Females of average age = 37). 41 to the real-words-nonce-context: 100% story condition and 40 to the control condition.

Procedure

Participants in each condition were initially tested on Wordsum and Pretest (following the procedure for Experiment 1) and then randomly assigned to either the nonsense-story condition or to a control group that was not exposed to a story. All participants were tested on the same set of drill-down questions.

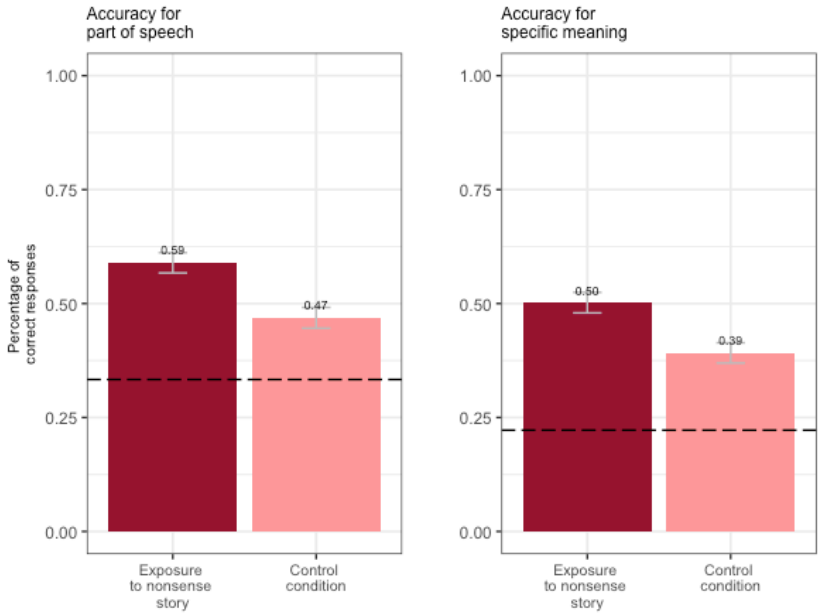


Figure 25. Group performance for Experiment 3. Horizontal dashed lines indicate chance-level. Error bars indicate +/-1SE of the mean.

Results and Discussion

Results are shown in Fig. 25. We found a significant effect of exposure to the linguistic context in inferring the part of speech [z = 3, SE = 0.18, p = .002] and the specific meaning [z = 2.6, SE = 0.2, p = .008] of real-rare words when compared with the control condition [accuracy_overall ~ condition+(1|subj id)] (Fig. 4). When examining only trials on which participants inferred the correct part of speech, the benefit of exposure to a nonce story on inferring the correct specific meaning was no longer significant [z = 0.5, SE = 0.31, p = .614] [accuracy_part of speech ~ condition+(1|subj id)+(1|word)]. To determine if the nonce-word context in the present study was more effective than in the equivalent condition of Experiment 2, we examined the interaction between experiment (Exp. 2 vs. Exp. 3) and condition (control group vs. real-words-nonce-context: 100%) [accuracy ~ condition*experiment+(1|subj id)+(1|word)]. This interaction was significant for both part-of-speech [z = 2.9, SE = 0.18, p = .003] and specific meaning [z = 2.5, SE = 0.21, p = .003]. As in Experiment 2, we found a significant effect of both frequency and salience of the target words. The benefit of being exposed to the nonce story was greater for more frequent words [for part of speech: z = 2.1, SE = 0.14, p = .035; for specific meaning: z = 2.5, SE = 0.15, p = .012] and more salient words [for part of speech: z = 2.2, SE = 0.14, p = .031; for specific

meaning: z = 2.6, SE = 0.16, p = .009]. Controlling for pretest scores, greater vocabulary knowledge (Wordsum) predicted higher accuracy for both part of speech [z = 3.4, SE = 0.091, p = .001] and specific meaning [z = 3.1, SE = 0.1, p = .002]. In Experiment 3, we observed an effect of context similar to that observed in Experiment 1. People were able to learn something about unknown words (here, real-rare words) from contexts in which all content words were replaced by nonce words. The only difference between Experiment 2 and the present study was in how word knowledge was assessed. A plausible conclusion (though in need of further testing) is that explicitly asking participants about a word's part of speech helps them deploy a more informative prior within which to consider the specific meanings.

General discussion

What are the limits of learning word meanings from language? Our results show that participants were able to learn something about what a word means from brief exposures to such seemingly meaningless contexts as "The stronk rourthed daft to a dweave luk as the slom zeuded rhiecing." For example, after reading a 600+ word nonsense-word story containing sentences like the one just used, 38% of participants correctly inferred that "stronk" means "an insect" compared to 0% of participants who were not exposed to the story. Our attempt to replicate and extend the results of Experiment 1 to using sparse contexts to inform the meanings of real, but rare and generally not known words (e.g., auscult, mitius), revealed that while partially informative contexts (40-60% of content words replaced with nonce words) were helpful, more sparse contexts (90%-100% of content words replaced) were not. We hypothesized that a key difference was the way that word knowledge was tested. In Experiment 1, participants' word knowledge was tested using a drill-down format that asked participants to first consider the word's part of speech. In Experiment 2, participants were asked to choose from among all the nine options visible at the same time making part of speech a less salient dimension of the word's meaning in the testing phase. In Experiment 3 we used the methods and materials of Experiment 2 with the test format of Experiment 1. Highlighting part-of-speech information using "drill-down" questions once again revealed that participants were able to use the nonce-story context to infer meanings of novel words. Experiments 2-3 further showed that the effects of context were positively associated with frequency and salience. Words that were more frequent and more salient benefited more from

context. While frequency is perhaps the most often used predictor in studies of word learning, to our knowledge, we are the first to examine the role of salience, defined here as the likelihood that people recall reading the word (see Table 1). What precisely makes a word salient requires future research. What information did people use to infer word meanings? In the all-nonce-word conditions of Experiments 1 and 3, greater than baseline performance cannot be explained by reliance on the meaning of English content words because no recognizable content words were present. There were three remaining sources of information: closed-class words, inflectional cues, and syntactic cues. Consider one sentence from the story: "He thecked up into a dweave luk to fruth in for a sparf snurv." The remaining pronouns and prepositions combined with inflectional cues can clearly be used to infer that, e.g., "thecked" is an action being performed by an animate agent and that a "dweave" is likely to be some kind of place. Implicit knowledge of English syntax such that verbs follow "to" and objects tend to come after verbs offers further guidance. What is remarkable is that participants are making these inferences in parallel across dozens or even hundreds of words and that a single exposure to the story is sufficient to achieve above baseline accuracies. Our work has two main limitations. First, successful use of sparse contexts involving mostly or exclusively nonce words clearly requires participants to already have sophisticated knowledge of English and so while it can help us understand how adults learn new words from context, it does not tell us how people learn enough English to make use of such sparse contexts. Second, present experiments do not tell us the relative importance of closed-class words, syntax, and morphology. Answering this question would require manipulating these sources of information independently. We can also gain additional insights by conducting studies such as this in more morphologically rich languages. As has been long known, people are able to learn something about a word's meaning from encountering it in context. What is surprising is just how sparse and seemingly uninformative that context can be. The facility that people show in inferring word meanings from such sparse and seemingly meaningless contexts suggests that we may be underestimating the role that morpho-syntactic and distributional cues have on both learning word meaning and on acquiring semantic knowledge that is embedded in language.

6.4. What do we learn from language? Summary and brief debate

If looked at within the bigger picture outlined in this chapter, the present study only targets one possible, maybe marginal, facet of the information provided us by the language system and it is indeed still far from addressing what would be the core aspect of our proposal, i.e. the role of the structural web of relationships in shaping our cognitive representations of meaning. We have already reviewed some of the general limitations in experimentally targeting this aspect of our proposal, at least at the current stage of our research, and some of the more specific limitations of the study in question. In this brief conclusion, we want to outline some of the positive aspects of our proposal in light of the current stage of the debate and of possible future lines of inquiry.

One first main consideration evaluates our theoretical proposal in regard to the linguistic and cognitive facts it allows to explain. Much has already been said on this subject in the previous sections. To summarize, our proposal waves together insights from theoretical linguistics about the nature and organization of meaning in language with a usage-based approach, which is now at the core of much of the research enterprise in the cognitive sciences. The structural claims about the organization of meaning are considered as the theoretical foundation of distributional approaches in computational linguistics, which, in turn, provide the methods and implementation to represent and computationally extract the web of relationships underlying the construction of meaning in language. The fundamental intake is that distributional semantics is in line with the Saussurian view of the language system and that, based on these grounds, distributional representations allow us to tap onto the "language-specific" level of linguistic meaning. Needless to say, we are extremely simplifying the problematic aspects of such claim, that have although been extensively discussed throughout the chapter. To the extent that distributional representations allow us to represent the intra-linguistic relationships between the units of language, distributional semantics can be considered in line with symbolic approaches to meaning.

However, as we have seen, this does not exhaust the discussion on what meaning is and how we represent it, but it rather gives us an insight on one specific level of analysis. To even attempt to give a somewhat exhaustive view on meaning, structural-distributional representations need to be understood in the broader framework of their role in proving us with relevant and specific information in the way we construct meaning at the cognitive level, which is essentially taken to be grounded and embodied. Similarly, the intersection of distributional and embodied

approaches sheds light onto the fact that linguistic meaning and the language system depend on our non-linguistic construction of the world as well as contribute in constructing it. Multimodality, in the sense of the intertwined nature of linguistic and non-linguistic aspects of our knowledge about the world, is the core element of our representations at both the linguistic and the cognitive level. The integration of structural-distributional and embodied theories of meaning (or at least the attempt to show where they converge, where they diverge, and to what extent they are complementary) moves us away from a dichotomic view of language (and cognition) as either symbolic or embodied and towards a unification of the two. Disentangling what kind of experience provides us with what kind of information in the way we build meaning is now at the center of a wide variety of theoretical, experimental, and modeling enterprises. Both claims about the multimodality of semantic representation need a much deeper investigation. In the first place, we are not fully discussing the role of experience and conceptual structure (or even how this structure might come to be or look like). Of course, a full-fledge account of multimodality would need to investigate both sides equally well, which in turns leads to the need to come up with methods that would eventually allow us to extract the specificities of both respectively. To this regard, our proposal addresses just one of the two core aspects that we have been identifying, i.e. the structure of language and DS as a way to tap onto it, and we are clearly leaving underspecified here what suggestions and methods could account for the more specific conceptual contents or the conceptual structures.

The specificity of our proposal is the focus on the structural aspects of meaning. Not only language might provide us with information that could not be available otherwise, but it also seems to present us with a way in which our experience is organized. We attempted to show how our proposal helps us making coherent sense of a wide variety of data from both the psycholinguistic and the computational pool. Our proposal intersects a linguistic definition of meaning with a conceptual definition of meaning, suggesting that the system of language, although maintaining a specificity at a certain level, nevertheless depends on our cognitive abilities and on the bodily and perceptual way in which we experience the world. By being an inherently usage-based approach in the sense described above, our proposal allows us to define a bottom-up perspective on the role of language in shaping cognition. Potentially useful insights may come from further developments at the intersection of structural-distributional and constructionist approaches to meaning, extending the discussion from the analysis of the meaning of word units (i.e. from lexical

semantics) to the meaning of construction at broader levels of analysis, such as sentences, argument structures and so on. Hence, constructionist and distributional approaches are not in contrast but could indeed be integrated to widen our understanding of how meaning is constructed at the language level as well as at the cognitive one. We should always keep in the back of our minds that semantic and syntactic, as well as pragmatic, aspects of meaning are not independent neither at the level of linguistic description nor at the level of establishing the role of language in shaping our mental representations. For instance, this seems to be one of the traits (and limitations) of our study, which, although providing us with interesting insights about the role of the structural aspects of the linguistic context in constructing the meaning of unknown words, it does not yet allow us to specify what aspects of context provide what kind of information. Future research will aim at designing methods to collect more specific evidence to this regard.

Hence, one strength of our proposal is that, while still being largely speculative in nature, it is inherently committed to outlining a unitary framework for the understanding of meaning in language and cognition. In this sense, our proposal just brings together in a coherent way some of the most prominent insights of the linguistic and the cognitive literature on meaning, and it hands back a theoretical perspective by means of which different aspects of meaning can be encompassed.

Notwithstanding, we are well aware that, throughout our discussion, we have barely mentioned two other perspectives that are fundamental for the study of meaning and are engrained in its very definition: neuroscientific approaches and the social and cultural aspects that are indeed constantly involved in how we behave in the world and understand reality. A proper account of our proposal in the light of neuroscientific and cultural-linguistic studies would require a parallel, and indeed as voluminous, work. The selection we made of the arguments is mostly due to our main intention to bring together insights from the Saussurian structural perspective with the distributional approach and the main topic of (second generation) cognitive accounts of meaning. We know that much has been left out and that much more could indeed be said about each and every of the points, theories, and arguments discussed so far.

In the more general conclusion that is about to follow, we will further address how our proposal allows us to understand and re-frame the idea of a cognitive function of language and of how it supports findings and proposals in the linguistic relativity tradition.

General Conclusions

Throughout this work we have talked about the relationship between language and thought in many ways and under very different perspectives. In the first place (see chapter 1), we tried to reconstruct (a part of) the history of the debate on the possible influence of language on thought. In the second place, we tried to highlight that within the relativistic debate much seems to be taken for granted in the notion of linguistic meaning that implicitly lies at the heart of many of the proposals we reviewed. We also tried to show how the mechanism of influence that is generally hypothesized to take place in the interaction between language and thought leaves many aspects of this relation underspecified and it seems to fail in explaining a good deal of linguistic intuitions as well as empirical data. But, most importantly, the relativistic discussion often assumes the possibility of separating language from thought without further discussing how and where this distinction may indeed be evident, or even possible. We tried to show how much of the debate in the cognitive sciences – and in those linguistic approaches that stem from the first as well as second cognitive revolution – hypothesize a very different situation from the one pictured by relativistic approaches. This is true especially when it comes to the definition of linguistic meaning. As we have seen many times, language is taken to merely mirror the structure and content of thought. We have extensively reviewed the shortcomings of such a view – first and foremost how language diversities (when taken seriously) lead to very counterintuitive claims in the framework of a strict language-as-a-window-on-thought view. Furthermore, if - in line with relativistic approaches - we aim at uncovering the effects of language on thought, then language needs to be taken as (at least at some level) different from thought. At the same time, it is undeniable that language is a cognitive phenomenon that depends and is likely to be shaped by other cognitive abilities and cognitive contents. In a nutshell, on the one hand, a neat distinction between language and thought is not as straightforward as relativistic accounts make it seem like and, on the other hand, language is not just a conventional set of labels that allows us to communicate however-else structured thoughts.

Consequently, the question we tried to address was: how can we maintain a language-specific level of linguistic meaning without giving up on a cognitive perspective? In chapter 2, we proposed to start by looking at the level of the language-internal systemic organization of meaning in languages. Our discussion took the lead from the Saussurian perspective as reinterpreted in the

work of De Mauro, and especially from the centrality of the intertwined nature of the *langue* and *parole* interface. The systemic nature of linguistic meaning – and hence what we called the "language-specific" level of meaning – is never separated from the speaker as a cognitive, social, and cultural individual. Quite the contrary, the system of language should be taken as a standardization of the uses of words made by the people of a given speaking community, and hence as never completely independent from the rest of our experience. It is by means of this always ongoing and dynamic process of standardization that language becomes a categorization – or, as De Mauro called it, a *pertinentization* – device that organizes the matter of our experience in a language-dependent way. The Saussurian perspective offers us a way to look at a "language-specific" level of linguistic meaning – that of the *langue* – as an abstraction from the multiplicity of the uses we make of language, and hence as a product of the speaker as a biologically, cognitive, and socially determined individual. This perspective allows us to translate the language and thought relationship onto a different level: language carves up the space of our experiences and thoughts in a language specific way, as much as it inherently depends on our biological-cognitive constraints. Hence, language and thought are not completely independent nor the first is merely a copy of the latter. The level of the systematic pertinentization operated by language offers us the ground to start evaluating the possibility of a "language-specific" level of meaning within the framework of a cognitive approach to language.

The obvious subsequent question is how the structure of language may look like, how we could analyze it, and what intuition about the influence of language on thought we could draw from such level. In chapter 3 we extensively reviewed some of the most established and well-known perspectives on the structure of language and on the structure of linguistic (or conceptual and linguistic) categories. In chapter 4 we turned to a more specific issue, i.e. how language should be organized for the purposes on language comprehension and production. We focused on the properties of the so-called *mental lexicon* as a *dictionary* and as a *network*. We tried to show how distributed and connectionist accounts of language learning and processing come to the conclusion that a mental storage of word (or sub-word) units may not be mandatory for establishing the mechanisms underlying our linguistic abilities. Words might as well work as activators of multiple webs of connections of representations in our mind that are not inherently linguistic, functioning as cues to meaning (understood as, generally speaking, mental content). However, what remains considerably underspecified in the word-as-cues to meaning approaches is the nature of linguistic

meaning from both a cognitive as well as a linguistic perspective. It could be argued that in order to clarify the word-as-cues perspective we would need a move similar to that operated in this work; i.e. we would need to identify the nature and organization of meaning in language and, subsequently, to theorize how the activation of the conceptual content or, more specifically, of the neural substrate may actually occur. The last aspect of the organization of language "in the mind" we considered was the data available from free word associations and the kind of structure they suggest. Free word associations are inevitably linked to an encyclopedic view of meaning in which the connections entailed between elicited words may be of various different sources, picturing a very wide variety of links that may hold together our memory for semantic content well beyond the connections provided by language itself. Needless to say, free word associations are not a direct way to tackle onto the organization of the "language-specific" level of meaning. Words are mostly taken as labels to identify and study a different level of organization, namely the cognitive organization of meaning. As we have seen, the idea that words can be viewed as placeholders for thoughts gives rise to a set of problems that ought to find a possible resolution in viewing language as something that is (at least at some level) autonomous from thought. We proposed that this level can be tackled by looking at how words behave in language and, especially, in the use we make of language.

The heart of our proposal is that the Saussurian relational notion of linguistic meaning as a dynamic process of meaning creation at the intersection of *langue* and *parole* can be operationalized in the framework of distributional semantic approaches. Chapter 5 is mostly dedicated to this topic. We tried to show how both in its historical roots as well as in the current methods and results, distributional accounts of linguistic meaning are inherently based on a relational view of meaning and allow us to quantify and represent its structural aspects. Furthermore, distributional semantics is inherently usage-based. The "language specific" level of the structure of language emerges from the use we make of language itself as a stabilization of senses and acceptations into structured webs of meaning relationships. The perspective put forward by distributional semantics seems to us as inherently in line with a Saussurian-Demaurian view on the nature and structure of linguistic meaning.

The operationalization the notion of meaning in terms of distributional analyses operated on linguistic corpora allows us to reflect on how the structure of meaning – which in this sense can be viewed as something in connection with but not reducible to other aspects of cognition – may

play a role in shaping the content and structure of our thought. Current researches in psychology and psycholinguistics show how meaning (both in language and in cognition) is inherently multimodal. The fact that the structure of language arises from the continuous interaction of the level of the system – *langue* – and of the individual uses – *parole* – leads us to idea that non-specifically linguistic content is always implicated and injected into our communicative practices while, at the same time, being inherently limited by the structure languages provide us with. On the other side, the way in which we learn meaning – i.e. the way in which our experiences become meaningful to us – is always shaped by a variety of factors, one of which is our experience with language itself. Our hypothesis is that the language we speak provides us with a "language-specific" level of organization of our experience from which we learn both structure and content. Obviously, we do not simply internalize this type of linguistic experience as it is. We seem to combine our linguistically-structured experience with other types of experiences – e.g. perceptual experience, and so on – when we progressively build our (internal) organization of meaning. The next (and, for now, last) step was then to come up with testable hypotheses on the role of the structure of language – as we built up to it throughout our work – as a relevant ingredient in the formation of our thoughts.

We started by looking at how the structure of language may influence the way we learn novel word meaning. Participants in our studies showed to be particularly apt at learning novel word meaning from linguistic context in which only morpho-syntactic and distributional cues where available. Our results suggest a possible interesting line of research on how the structural aspects of meaning may contribute in the formation and the shape of our knowledge of the world. As we have already pointed out many times, this is just a first step into a yet broadly unexplored line of research, in which the focus is on how the relational aspects of meaning may contribute in shaping what we know and how we come to know it.

How is our approach different from the more traditional relativist claim that *the way we speak shapes the way we think*? Bottom line, we are still claiming that the way we speak shapes the way we think. However, as it might already be clear, the mechanism by which this comes to be the case is essentially reversed. Moreover, we take a very different perspective on how language and thought can be disentangled. Let us remark this latter point first. On the one side, we argued that language and thought *cannot* be easily disentangled: we do not seem to have two separate sets of representations, i.e. one mental storage for concepts and one for words, that then interfere with

each other giving rise to linguistically-shaped non-linguistic behaviors. To the extreme, we might not have a mental storage for words at all. The content and structure of meaning is always cognitive and linguistic in the sense that the more strictly linguistic aspects of meaning cooperate in giving rise to our mental representations, and the non-linguistic content is always implicated in the way we create meaning in language. In the framework of this intertwined nature of language and thought, our proposal aims at delineating the possibility of identifying a "language-specific" level of meaning in language. As we have pointed out multiple times already, language exhibits a structure and organization specific to language that emerges from the uses we make of language in the continuous interaction between idiosyncratic contents and the already established, socially shared, structure of language. To sum things up, although language and thought cannot be separated, language exhibits a structure that is not completely equivalent with the structure of thought. The relational structure of language is what we should look at when evaluating the specificity of language. Furthermore, the "language-specific" structure of language might be informative of the way in which language is implicated in shaping our cognitive content. In this sense, our perspective is reversed compared to previous relativistic approaches. Language is one of the ingredients that gives rise to our cognitive content in the first place. When acquiring language, we are exposed to a structured, language-specific, experience that becomes part of the way in which we experience the world and create meaning.

Let us conclude with some limitations of our proposal and with some outstanding open questions. It is rather clear that our proposal of identifying the "language-specific" level of linguistic meaning in the system of language as envisaged by Saussure ought to be corroborated by research on the cognitive plausibility of the mechanisms underlying the very formation of languages. For instance, we could ask whether mechanisms of language formation and language change that can be hypothesized starting from a Saussurian perspective – such as the *arbitrariness* of language – are actually cognitive plausible and if they might respond to specific evolutionary selective pressures. The main point of this criticisms is that the idea of a "language-specific" level of meaning should make sense in terms of both an evolutionary and developmental perspectives. There should be cognitive and evolutionary reasons for languages to be as different as they are and to exhibit the kind of specificity we outlined. Our proposal did not address this point. However, some suggestions can be made to this regard.

It might well be the case that a "language-specific" structure of language makes languages easier to learn. In fact, all natural languages must be learnable. If they were not, they could not be propagated to subsequent generations of learners (Christiansen & Chater, 2008; Kirby, Cornish, & Smith, 2008). We might then ask to what extent the design features that we observe in contemporary languages reflect a selection pressure on language to be maximally learnable (Bentz, Alikaniotis, Cysouw, & Ferrer-i-Cancho, 2017; Lupyan & Christiansen, 2002; B. Thompson, Kirby, & Smith, 2016). In the project we presented in chapter 6, we showed how people may learn many word meanings not from associating the word with external referents, but from language itself. In a subsequent experiment, we examined how syntax, morphology, and function words contribute to creating language in a way that makes such learning possible.

What might explain the remarkable performance we observed in our first study? One might seek an answer in human cognition, e.g., in pragmatic inference (Frank & Goodman, 2014). A very different (though not mutually exclusive) kind of explanation is that it is the structure of language itself that allows people to learn meanings from language itself. Our second project focuses on this second explanation by using neural network models (which lack all pragmatic abilities) trained on natural language. By training on augmented versions of language, we can learn whether removing certain linguistic devices (e.g., function words, morphology) impacts how well the models learn the sort of linguistic structure that allows people to use highly impoverished contexts to successfully infer word meanings. We trained recurrent (Long short-term memory) neural networks (Gulordava, Bojanowski, Grave, Linzen, & Baroni, 2018) on 5 versions of English Wikipedia: (1) Original, (2) Shuffled (all words within each sentence shuffled), (3) Partial Shuffling (20% of words within each sentence shuffled), (4) Function-word shuffling (only function words shuffled within each sentence), (5) Lemmatized (Inflected content words replaced with their lemmas). Each model then assigned a probability of encountering words produced by human subjects in a separate sentence prediction task. For example, given the sentence "Instead he fell right back into the ___", typical human responses include "mud", "pool", and "hole". By giving the same sentence as an input to each trained model, we can estimate how likely it is to predict these words, a measure of how well it has learned the language in each condition.

Shuffling words within each sentence had a devastating impact on what the models were able to learn (though the models still benefited from the maintained distribution of word frequencies). Shuffling function words and lemmatization led to similar, significant, decrements

in performance. Surprisingly, partial shuffling had no effect (indeed on some measures partial shuffling improved performance, hinting that increased word order variability can be beneficial). Taken together, our simulations show that morphology and function words (despite often being communicatively redundant) are critically important for bootstrapping learning of language structure. The results of these brief simulations suggest that some aspects of the structure of language (in this case the more syntactical and morphological ones) may indeed be evolutionary advantageous. Needless to say, this is just a first approximation of what might be interesting to look at in evolutionary and developmental terms to corroborate the idea that a "language-specific" level of meaning may indeed be selectively important. Further research and implementation of recurrent neural networks is needed to evaluate the possibility of applying the same experimental methods to other structural aspects of language closer to our theoretical perspective.

Finally, the applicability of our proposal in other areas of research needs further investigation. As we have already pointed out many times, this work should be understood as the theoretical grounds on which to build empirical researches both to corroborate our ideas as well as to open new lines of inquiry. The possible range of applications of a successful integration of the relational view of meaning into psycholinguistic approaches is wide and diverse.

We introduce here a project carried out in collaboration with Katie Hoemann and Lisa Feldman Barrett (Interaffective Science Labaoratory, Department of Psychology, Northeastern University Boston), started during a vising scholar research period from August to December 2018 and still on-going. The complete results and discussion are under submission to the international journal *Nature Human Behavior* and have been presented at the 15th International Cognitive Linguistics Conference (August, 2019) and the bi-annual meeting of the International Society for Research on Emotion (June, 2019). In this project, we have been investigating how, starting from the structure of the lexicon, we could come up with better measurements of cross-linguistic translatability (i.e. semantic alignment) – focusing mostly of the translatability of supposedly untranslatable emotion words - by integrating intuitions from the proposal discussed in this work with methodological techniques from networks analysis. To this regard, purportedly 'untranslatable' emotion words pose interesting questions that can perhaps find an interesting answer within the proposal discussed in this work. For example, Tagalog speakers have a word, "gigil", for "when something is so cute that you want to squeeze it"; second language learners of Dutch have a difficult time appreciating the nuances of "gezelligheid"; and Italian does not have

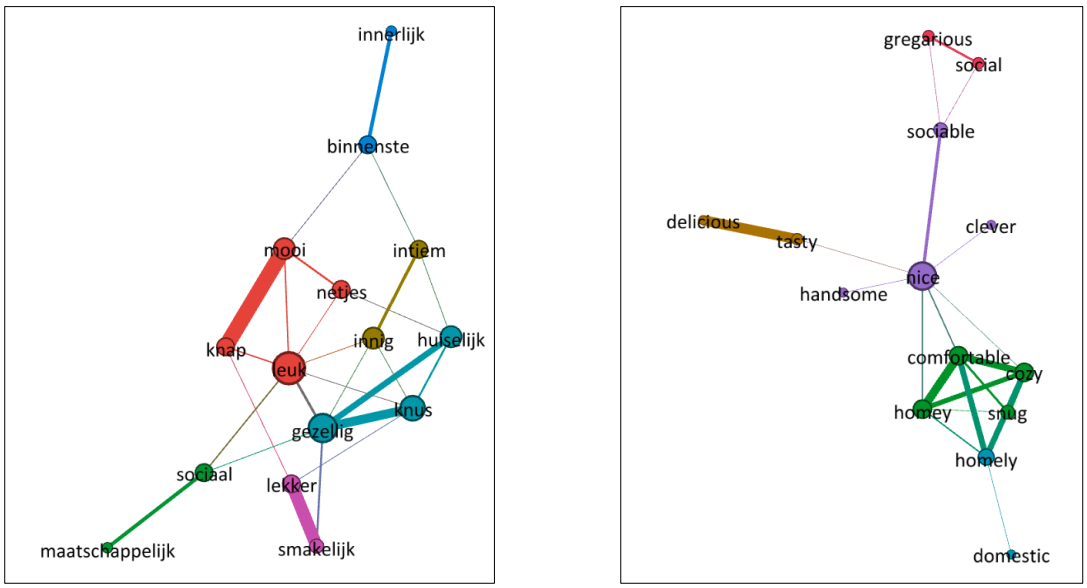


Figure 26, Networks constructed using free association data: a) Dutch "gezellig"; b) possible English translations. Semantic spaces generated via back translation to produce equivalent number of words. Edges weighted based on forward association strength; nodes sized according to number of connections. Node and edge color reflect community assignment, illustrating distinct aspects of meaning.

an easy translation for English "excitement". But are these words actually untranslatable, or does translatability perhaps exist on a continuum? Previous work has employed a variety of methods for comparing word meaning across languages (see Ogarkova, 2016 for a review), including property generation or feature rating, free association, scenario elicitation, metaphor analysis, and labeling or naming. However, most of these methods are focused on describing the nature of differences in meaning (e.g., based on conceptual features), or comparing entire semantic fields, rather than measuring the similarity or dissimilarity of individual words.

In this project, we use free association data from the Small World of Words study (De Deyne, Navarro, & Storms, 2013; De Deyne & Storms, 2008b) to construct semantic networks for 'untranslatable' Dutch words (e.g., "gezellig") and their possible English translations. Network analyses of free association data have been used to model lexical knowledge and semantic growth (Steyvers & Tenenbaum, 2005), but have not yet been used to compare meaning across languages. Moreover, by using closely-related languages, we are able generate networks that are maximally comparable (Majid, Jordan, & Dunn, 2015). We populated semantic spaces for Dutch and English target words using the well-known back translation protocol. We then characterized these semantic spaces in terms of their network metrics. For each Dutch to English comparison, we selected

networks with equivalent metrics and evaluated the similarity of target words' behavior within their respective networks using these same metrics (e.g., degree, betweenness centrality, clustering coefficient). Using this data-driven approach, we found no (single) English word matches for Dutch targets (figure 26).

Semantic networks can also be constructed based on distributional semantic models (e.g. Utsumi, 2015). Recently, DSMs have been used to compare semantic fields across languages (B. Thompson et al., 2018), and to generate word-level predictors of second language learning (Hopman, Thompson, Austerweil, & Lupyan, 2018). To contextualize our results from the free association networks, we presented DSM networks based on lexical neighborhood data from Continuous Bag-of-Words (CBOW) models trained for Dutch and English, which have been shown to accurately predict behavioral data in both languages (Mandera et al., 2017). These DSM networks allow us to compare words based on both paradigmatic and syntagmatic relationships they maintain within their respective languages, thereby providing insight into aspects of word meaning that may be ignored by other methods. Taken together, our results illustrate a novel means of assessing translatability. Furthermore, this project has now seen a rapid growth thanks to the integration of more specific network techniques in the theoretical framework we just presented. We are now moving in the direction of mapping the whole structure of the lexicon and in creating cross-linguistically aligned networks. Very briefly, comparisons between words in the method we just outlined as still indirect and heavily depend on the initial choices made by researchers. In our new method, we bring languages together into the same 'semantic space' making comparisons between words direct by means of co-registered networks (Hoemann, K., De Luca, M., Henry, T., & Barrett, L. F., in preparation).

This last project shows that our proposal of reinforcing the attention to the structural aspects of meaning may lead to interesting and innovating research paradigms with a broad spectrum of possible applications, including the development of linguistically-informed computational approaches to translatability issues. Furthermore, this project highlights that methods based on the analysis of the internal structure of language may take advantage of a variety of data, from distributional semantics to free word associations. Allow us one final remark. The comparison between DSM data and free word association data is still a matter of debate and, as we have shown, broad consensus among researchers is yet not available on the kind of meaning represented by the two sets of data. Nonetheless, as we have already mentioned, the two are not

mutually exclusive. On the contrary, developing a better understanding of what they represent and what kind of structure we might observe is one of our leading interests, inasmuch as it could allow us to shed better light on the content and structure of our representations both in cognition and in language.

References

Acquaviva, P., Lenci, A., Paradis, C., & Raffaelli, I. (forth). Models of lexical meaning. In V. Pirelli, I. Plag, & W. Dressler (Eds.), *Word knowledge and word usage: A cross-disciplinary guide to the mental lexicon* (p. forth). Berlin: de Gruyter Mouton.

Aitchison, J. (1987). *Words in the mind: An introduction to the mental lexicon*. Oxford, UK ; New York, NY, USA: Blackwell.

Albano Leoni, F. (2009). *Dei suoni e dei sensi. Il volto fonico delle parole*. Bologna: Il Mulino.

Alford, D. K. H. (1981). Is Whorf's Relativity Einstein's Relativity? *Proceedings of the Seventh Annual Meeting of the Berkeley Linguistics Society*, 13–26.

Alinei, M. (1974). *La Struttura del lessico*. Il Mulino.

Allport, D. A., & Funnell, E. (1981). Components of the Mental Lexicon. *Philosophical Transactions of the Royal Society B: Biological Sciences*, (1077), 397–410. https://doi.org/10.1098/rstb.1981.0148

Anderson, J. A. (1977). Neural models with cognitive implications. In D. LeBerge & S. J. Samuels (Eds.), *Basic processes in reading perception and comprehension* (pp. 27–90). Hillsdale, NJ: Erlbaum.

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling Embodied and Distributional Accounts of Meaning in Language. *Topics in Cognitive Science*, *6*(3), 359–370. https://doi.org/10.1111/tops.12096

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463–498. https://doi.org/10.1037/a0016261

Apresjan, Û. D. (1966). Analyse distributionnelle des significations et champs sémantiques structurés. *Langages*, *1*(1), 44–74. https://doi.org/10.3406/lgge.1966.2865

Baayen, H. R. (2007). Storage and Computation in the Mental Lexicon. In G. Jarema & G. Libben (Eds.), *The Mental Lexicon: Core Perspectives* (pp. 81–104). Amsterdam: Elsevier.

Bally, C. (1932). *Linguistique générale et linguistique française*. Berne: Francke.

Baronchelli, A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in Cognitive Science. *Trends in Cognitive Sciences*, *17*(7), 348–360. https://doi.org/10.1016/j.tics.2013.04.010

Baroni, M. (2016). Grounding Distributional Semantics in the Visual World. *Language and Linguistics Compass*, *10*(1), 3–13. https://doi.org/10.1111/lnc3.12170

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. Context-predicting semantic vectors. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 238–247. https://doi.org/10.3115/v1/P14-1023

Baroni, M., & Lenci, A. (2010). Distributional Memory: A General Framework for Corpus-Based Semantics. *Computational Linguistics*, *36*(4), 673–721. https://doi.org/10.1162/coli_a_00016

Baroni, M., & Lenci, A. (2011). How we BLESSed distributional semantic evaluation. *Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics*, 1–10. Stroudsburg, PA: Assoc. Comput. Linguist.

Barsalou, L. W. (1992). Frame, Concepts and Conceptual Fields. In A. Lehrer & E. F. Kittay (Eds.), *Frames, Fields and Contrasts* (pp. 21–74). Hillsdale, N.J: Lawrence Erlbaum Associates.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* (pp. 29–101). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577–609; discussion 610-660.

Barsalou, L. W. (2005). Continuity of the conceptual system across species. *Trends in Cognitive Sciences*, *9*, 309–311. https://doi.org/10.1016/j.tics.2005.05.003

Barsalou, L. W., & Prinz, J. J. (1997). Mundane creativity in perceptual symbol systems. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 267–307). https://doi.org/10.1037/10227-011

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and Embodiment: Debates on meaning and cognition* (pp. 245–283). New York: Oxford University Press.

Basile, G. (2001). *Le parole nella mente*. FrancoAngeli.

Basile, G. (2005). Può darsi una semantica senza pragmatica o viceversa? Cosa accade nel processo di denominazione. In A. Frigerio & S. Raynaud (Eds.), *Significare e comprendere. La semantica del linguaggio verbale*. Aracne.

Basile, G. (2010a). I percorsi del senso. In che modo il senso prende forma in parole. *Bollettino di italianistica*, *VII*(I), 9–29.

Basile, G. (2010b). Significazioni e rappresentazioni: Teoria di un percorso comune. *Teorie & Modelli*, *XV*(2–3), 73–88.

Basile, G. (2012). *La conquista delle parole. Per una storia naturale della denominazione*. Roma: Carocci.

Basile, G. (2016). Towards a Pragmatic-Semantic Continuum. The process of naming. In K. Allan, A. Capone, & I. Kecskes (Eds.), *Pragmens and Theories of Language Use*. Springer.

Basile, G. (2019). Il lessico come sistema: Relationi semantiche e polisemia. In F. Casadei & G. Basile (Eds.), *Lessico ed educazione linguistica* (pp. 41–72). Roma: Carocci.

Beckage, N. M., & Colunga, E. (2015). Language Networks as Models of Cognition: Understanding Cognition through Language. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, & B. Job (Eds.), *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks* (Softcover reprint of the original 1st ed. 2016 edition, pp. 3–38). s.l: Springer.

Beckner, C., Blythe, R., Bybee, J. L., Christiansen, M. H., Croft, W., Ellis, N. C., … Schoenemann, T. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning*, *59*, 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x

Bentz, C., Alikaniotis, D., Cysouw, M., & Ferrer-i-Cancho, R. (2017). The Entropy of Words—Learnability and Expressivity across More than 1000 Languages. *Entropy*, *19*(6), 275. https://doi.org/10.3390/e19060275

Bergen, B. K. (2012). *Louder than words: The new science of how the mind makes meaning* (Vol. 50). Retrieved from http://choicereviews.org/review/10.5860/CHOICE.50-5297

Bergen, B. K., & Chang, N. (2013). Embodied Construction Grammar. *The Oxford Handbook of Construction Grammar*. https://doi.org/10.1093/oxfordhb/9780195396683.013.0010
Bergen, B. K., & Chang, N. C. (2005). Embodied construction grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive grounding and theoretical extensions* (pp. 147–190). Amsterdam & Philadelphia: John Benjamins Publishing Company.
Berlin, B. (1992). *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies* (1st Edition edition). Princeton, N.J: Princeton University Press.
Berlin, B., & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley and Los Angeles: University of California Press.
Bhatia, S. (2017). The semantic representation of prejudice and stereotypes. *Cognition*, *164*, 46–60. https://doi.org/10.1016/j.cognition.2017.03.016
Bhatia, S., Goodwin, G. P., & Walasek, L. (2018). Trait Associations for Hillary Clinton and Donald Trump in News Media: A Computational Analysis. *Social Psychological and Personality Science*, *9*(2), 123–130. https://doi.org/10.1177/1948550617751584
Black, M. (1959). Linguistic Relativity: The Views of Benjamin Lee Whorf. *The Philosophical Review*, *68*(2), 228. https://doi.org/10.2307/2182168
Bloomfield, L. (1933). *Language*. University of Chicago Press.
Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4–5), 175–308. https://doi.org/10.1016/j.physrep.2005.10.009
Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Place.
Boleda, G. (2020). Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics*, Accepted. https://doi.org/10.1146/annurev-linguistics-011619-030303
Boole, G. (1854). *The laws of thought (1854)*. The Open court publishing company.
Borge-Holthoefer, J., & Arenas, A. (2010). Semantic Networks: Structure and Dynamics. *Entropy*, *12*(5), 1264–1302. https://doi.org/10.3390/e12051264
Boroditsky, L. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748
Boroditsky, L. (2011). How Language Shapes Thought. *Scientific American*, *304*(2), 62–65. https://doi.org/10.1038/scientificamerican0211-62
Boroditsky, L., & Prinz, J. (2008). What thoughts are made of. In G. R. Semin & E. R. Smith (Eds.), *Embodied Grounding: Social, Cognitive, Affective, and Neuroscientific Approaches* (p. 19). Cambridge University Press.
Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, Syntax, and Semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 61–79). MIT Press.
Bowerman, M. (1973). *Early syntactic development: A cross-linguistic study with special reference to Finnish*. Cambridge: Cambridge University Press.
Bowerman, M. (1996a). Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 145–176). Retrieved from https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_1456708

Bowerman, M. (1996b). The origins of children's spatial semantic categories: Cognitive versus linguistic determinants. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 145–176). Cambridge University Press.

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). https://doi.org/10.1017/CBO9780511620669.018

Bowerman, M., & Levinson, S. C. (Eds.). (2001). *Language Acquisition and Conceptual Development*. Cambridge, UK ; New York: Cambridge University Press.

Broccias, C. (2013). Cognitive Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar*. OUP USA.

Brown, C. M. (1993). Factors affecting the acquisition of vocabulary: Frequency and saliency of words. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 263– 286). Norwood, NJ: Ablex.

Brown, R. (1973). *A First Language* (Reprint 2013 ed. edition). Harvard University Press.

Brown, R. (1976). Reference in Memorial Tribute to Eric Lenneberg. *Cognition*, *4*(2), 125–153. https://doi.org/10.1016/0010-0277(76)90001-9

Brown, R., & McNeill, D. (1966). The "Tip of the Tongue" Phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*, 325–337.

Brown, R. W. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, *55*(1), 1–5. https://doi.org/10.1037/h0041199

Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology*, *49*(3), 454–462. https://doi.org/10.1037/h0057814

Browne, B. (1989). Effects of Vocabulary Difficulty and Text Length on Word Definition and Prose Recall. *The Journal of General Psychology*, *116*(4), 385–392. https://doi.org/10.1080/00221309.1989.9921125

Bruni, E., Boleda, G., Baroni, M., & Tran, N.-K. (2012). Distributional Semantics in Technicolor. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, 136–145. Retrieved from http://dl.acm.org/citation.cfm?id=2390524.2390544

Bruni, E., Tran, G. B., & Baroni, M. (2011). Distributional Semantics from Text and Images. *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, 22–32. Retrieved from http://dl.acm.org/citation.cfm?id=2140490.2140493

Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal Distributional Semantics. *Journal of Artificial Intelligence Research*. https://doi.org/10.1613/jair.4135

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526. https://doi.org/10.3758/BF03193020

Bullinaria, J. A., & Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: Stop-lists, stemming, and SVD. *Behavior Research Methods*, *44*(3), 890–907. https://doi.org/10.3758/s13428-011-0183-8

Bybee, J. L. (1985). *Morphology: A Study of the Relation Between Meaning and Form*. John Benjamins Publishing.

Bybee, J. L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, *82*(4), 711–733. https://doi.org/10.1353/lan.2006.0186

Bybee, J. L. (2010). *Language, Usage and Cognition*. Cambridge University Press.

Bybee, J. L., & Beckner, C. (2015). Usage-Based Theory. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis*. https://doi.org/10.1093/oxfordhb/9780199677078.013.0032

Bybee, J. L., & Hopper, P. J. (2001). *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing.

Bybee, J. L., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, *22*(2–4), 381–410. https://doi.org/10.1515/tlir.2005.22.2-4.381

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Camacho-Collados, J., & Pilehvar, M. T. (2018). From Word To Sense Embeddings: A Survey on Vector Representations of Meaning. *Journal of Artificial Intelligence Research*, *63*, 743–788. https://doi.org/10.1613/jair.1.11259

Cañas, J. J. (1990). Associative Strength Effects in the Lexical Decision Task. *The Quarterly Journal of Experimental Psychology Section A*, *42*(1), 121–145. https://doi.org/10.1080/14640749008401211

Capitán, J. A., Borge-Holthoefer, J., Gómez, S., Martinez-Romo, J., Araujo, L., Cuesta, J. A., & Arenas, A. (2012). Local-Based Semantic Navigation on a Networked Representation of Information. *PLoS ONE*, *7*(8), e43694. https://doi.org/10.1371/journal.pone.0043694

Caramazza, A. (1997). How Many Levels of Processing Are There in Lexical Access? *Cognitive Neuropsychology*, *14*(1), 177–208. https://doi.org/10.1080/026432997381664

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(06). https://doi.org/10.1017/S0140525X02000122

Casadei, F. (1999). Significato ed esperienza. Linguaggio, cognizione e realtà. In *Semantica. Teorie, tendenze e problemi contemporanei* (pp. 79–116). Roma: Carocci.

Casadei, F. (2003a). *Lessico e semantica*. Roma: Carocci.

Casadei, F. (2003b). Per un bilancio della semantica cognitiva. In L. Gaeta & S. Luraghi (Eds.), *Introduzione alla linguistica cognitiva* (pp. 37–56). Roma: Carocci.

Casasanto, D. (2008). Who's Afraid of the Big Bad Whorf? Crosslinguistic Differences in Temporal Language and Thought. *Language Learning*, 17.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, Mass: MIT Press.

Chen, X., Liu, Z., & Sun, M. (2014). A Unified Model for Word Sense Representation and Disambiguation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1025–1035. https://doi.org/10.3115/v1/D14-1110

Chomsky, N. (1957). *Syntactic Structures*. Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Christiansen, M. H. (2018). Implicit Statistical Learning: A Tale of Two Literatures. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12332

Christiansen, M. H., & Chater, N. (1999). Connectionist Natural Language Processing: The State of the Art. *Connectionist Natural Language Processing*, *23*(4), 417–437.

Christiansen, M. H., & Chater, N. (2001a). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences*, *5*(2), 82–88. https://doi.org/10.1016/S1364-6613(00)01600-4
Christiansen, M. H., & Chater, N. (2001b). Connectionist Psycholinguistics in Perspective. In M. H. Christiansen (Ed.), *Connectionist Psycholinguistics* (New edizione, pp. 19–75). Westport, Conn: Praeger Pub Text.
Christiansen, M. H., & Chater, N. (2001c). Connectionist Psycholinguistics: The Very Idea. In M. H. Christiansen & N. Chater (Eds.), *Connectionist Psycholinguistics* (New edizione, pp. 1–15). Westport, Conn: Praeger Pub Text.
Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*(5), 489–509. https://doi.org/10.1017/S0140525X08004998
Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (1994). *A Critique of Pure Vision'*. 25.
Chwilla, D. J., & Kolk, H. H. J. (2002). Three-step priming in lexical decision. *Memory & Cognition*, *30*(2), 217–225. https://doi.org/10.3758/BF03195282
Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again* (Reprint edizione). Cambridge, Mass.: Bradford Books.
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477
Clark, S. (2015). Vector Space Models of Lexical Meaning. In S. Lappin & C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory* (pp. 493–522). https://doi.org/10.1002/9781118882139.ch16
Collins, A. M., & Loftus, E. F. (1975). A Spreading-Activation Theory of Semantic Processing. *Psychological Review*, 407–428.
Collins, A. M., & Quillian, M. R. (1969). Retrieval Time from Semantic Memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247.
Coltheart, M. (2004). Are There Lexicons? *The Quarterly Journal of Experimental Psychology Section A*, *57*(7), 1153–1172. https://doi.org/10.1080/02724980443000007
Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204
Cong, J., & Liu, H. (2014). Approaching human language with complex networks. *Physics of Life Reviews*, *11*(4), 598–618. https://doi.org/10.1016/j.plrev.2014.04.004
Coseriu, E. (1958). *Sincronía, Diacronía e Historia*. Montevideo.
Coseriu, E. (1964). Pour une sémantique diachronique structurale. *Travaux de Linguistique et de Littérature*, *2*.
Coseriu, E. (1967). Lexicalische Solidaritäten. *Poetica*, *1*.
Coseriu, E. (1971a). Solidarietà lessicali. In R. Simone (Ed.), *Teoria del linguaggio e linguistica generale* (pp. 303–316). Bari: Laterza.
Coseriu, E. (1971b). *Teoria del linguaggio e linguistica generale, sette studi*. Roma-Bari: Laterza.
Coseriu, E. (1973). *Lezioni di linguistica generale*. Torino: Boringhieri.
Costa, L. da F., Oliveira Jr., O. N., Travieso, G., Rodrigues, F. A., Boas, P. R. V., Antiqueira, L., … da Rocha, L. E. C. (2011). Analyzing and Modeling Real-World Phenomena with

Complex Networks: A Survey of Applications. *Advances in Physics*, *60*(3), 329–412. https://doi.org/10.1080/00018732.2011.572452
Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press.
Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge University Press.
Curran, J. R. (2003). *From Distributional to Semantic Similarity*. University of Edinburgh - School of Informatics, Edinburgh, UK.
Cutler, A. (1983). Lexical complexity and sentence processing. In *The process of language understanding*. New York: Wiley.
Cutler, A., & Fay, D. A. (1982). One Mental Lexicon, Phonologically Arranged: Comments on Hurford's Comments. *Linguistic Inquiry*, *13*(1), 107–113.
Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a Stone-Age tribe. *Nature*, (398), 203–204.
De Deyne, S., Kenett, Y. N., Anaki, D., Faust, M., & Navarro, D. (2017). Large-scale network representations of semantics in the mental lexicon. In M. N. Jones (Ed.), *Frontiers of cognitive psychology. Big data in cognitive science* (pp. 174–202). New York, NY: Routledge/Taylor & Francis Group.
De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2018). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1115-7
De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–498. https://doi.org/10.3758/s13428-012-0260-7
De Deyne, S., & Storms, G. (2008a). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213–231. https://doi.org/10.3758/BRM.40.1.213
De Deyne, S., & Storms, G. (2008b). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods*, *40*(1), 198–205. https://doi.org/10.3758/BRM.40.1.198
De Deyne, S., & Storms, G. (2015). Word Associations. In J. R. Taylor (Ed.), *The Oxford Handbook of the Word*. https://doi.org/10.1093/oxfordhb/9780199641604.013.018
De Deyne, S., Verheyen, S., & Storms, G. (2015). Structure and Organization of the Mental Lexicon: A Network Approach Derived from Syntactic Dependency Relations and Word Associations. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, & B. Job (Eds.), *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks* (Softcover reprint of the original 1st ed. 2016 edition, pp. 47–79). s.l: Springer.
De Luca, M. (2017). Reexamining linguistic relativity: The approach and method of John A. Lucy. *BLITYRI*, *VI*(2), 119–140.
De Mauro, T. (1965). *Introduzione alla semantica*. Roma: Laterza.
De Mauro, T. (1967a). Introduzione e Commento. In F. de Saussure (Ed.), *Corso di linguistica generale*. Laterza.
De Mauro, T. (1967b). *Ludwig Wittgenstein. His Place in the Development of Semantics*. Dordrecht: Kluwer.
De Mauro, T. (1971). *Senso e significato: Studi di semantica teorica e storica*. Adriatica.
De Mauro, T. (1982a). *Minisemantica dei linguaggi non verbali e delle lingue* (1 edizione). Roma: Laterza.
De Mauro, T. (1982b). Semantica. In *Enciclopedia del Novecento*. Treccani.

De Mauro, T. (1994). *Capire le parole* (1 edizione). Roma: Laterza.
De Mauro, T. (2008). *Lezioni di linguistica teorica* (2 edizione). Roma: Laterza.
De Palo, M. (2003). Le semantiche postsaussuriane. *Rivista Italiana Di Linguistica e Dialetto-Logia*, *5*, 109–126.
De Palo, M. (2016). *Saussure e gli strutturalismi. Il soggetto parlante nel pensiero linguistico del Novecento*. Roma: Carocci.
de Vega, M., Graesser, A., & Glenberg, A. (2008). Reflecting on the debate. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 397–440). https://doi.org/10.1093/acprof:oso/9780199217274.001.0001
Deese, J. (1965). *The Structure of Associations in Language and Thought*. Johns Hopkins Press.
Dell, G. S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, *93*(3), 283–321.
Dell, G. S. (1990). Effects of Frequency and Vocabulary Type on Phonological Speech Errors. *Language and Cognitive Processes*, *5*(4), 313–349. https://doi.org/10.1080/01690969008407066
Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 611–629. https://doi.org/10.1016/S0022-5371(81)90202-4
Deutscher, G. (2010). *Through the Language Glass: Why the World Looks Different in Other Languages*. Metropolitan Books.
Devitt, M., & Sterelny, K. (1987). *Language and Reality*. Cambridge, Mass: MIT Press.
Deyne, S. D., Verheyen, S., Perfors, A., & Navarro, D. J. (2015). Evidence for widespread thematic structure in the mental lexicon. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, 6.
Diessel, H. (2015). Usage-based construction grammar. In E. Dabrowska & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 295–321). https://doi.org/10.1515/9783110292022-015
Diessel, H. (2017). Usage-based linguistics. In M. Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics*. New York: Oxford University Press.
Diodato, F. (2013). *Teorie semantiche. Dal segno al testo*. Napoli: Liguori.
Diodato, F. (2015). *Teorie della categorizzazione. Dal modello classico ai prototipi e oltre*. Napoli: Liguori.
Dove, G. (2011). On the need for Embodied and Dis-Embodied Cognition. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00242
Dove, G. (2016). Three symbol ungrounding problems: Abstract concepts and the future of embodied cognition. *Psychonomic Bulletin & Review*, *23*(4), 1109–1121. https://doi.org/10.3758/s13423-015-0825-4
Dummett, M. A. E. (1993). *The seas of language*. Clarendon Press.
Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Busseniu, Trans.). New York: Teachers College, Columbia University.
Eco, U. (1984). *Semiotica e filosofia del linguaggio*. Torino: Einaudi.
Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100. https://doi.org/10.1016/j.cognition.2015.06.008
Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.
Elman, J. L. (1991). Distributed Representations, Simple Recurrent Networks, And Grammatical Structure. *Machine Learning*, *7*, 195–225.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, *8*(7), 301–306. https://doi.org/10.1016/j.tics.2004.05.003
Elman, J. L. (2009). On the Meaning of Words and Dinosaur Bones: Lexical Knowledge Without a Lexicon. *Cognitive Science*, *33*(4), 547–582. https://doi.org/10.1111/j.1551-6709.2009.01023.x
Elman, J. L. (2011). Lexical knowledge without a lexicon? *The Mental Lexicon*, *6*(1), 1–33. https://doi.org/10.1075/ml.6.1.01elm
Elman, J. L., Bates, E. A., Johnson, M. H., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press.
Emmorey, K. D., & Fromkin, V. A. (1988). The Mental Lexicon. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey: Volume 3, Language: Psychological and Biological Aspects* (pp. 124–149). Cambridge University Press.
Erk, K. (2012). Vector Space Models of Word Meaning and Phrase Meaning: A Survey. *Language and Linguistics Compass*, *6*(10), 635–653. https://doi.org/10.1002/lnco.362
Erk, K. (2016). What do you know about an alligator when you know the company it keeps? *Semantics and Pragmatics*, *9*. https://doi.org/10.3765/sp.9.17
Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X
Evans, V. (2009). *How Words Mean: Lexical Concepts, Cognitive Models, and Meaning Construction*. Retrieved from https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199234660.001.0001/acprof-9780199234660
Evans, V. (2010). Figurative language understanding in LCCM Theory. *Cognitive Linguistics*, *2110*, 601–662. https://doi.org/10.1515/cogl.2010.020
Evans, V. (2013). *Language and Time: A Cognitive Linguistics Approach*. Cambridge University Press.
Evans, V. (2014). *The Language Myth: Why Language Is Not An Instinct*. Cambridge, United Kingdom ; New York: Cambridge University Press.
Evans, V. (2015a). *The Crucible of Language: How Language and Mind Create Meaning*. Cambridge, United Kingdom: Cambridge University Press.
Evans, V. (2015b). What's in a Concept? Analog versus Parametric Concepts in LCCM Theory. In E. Margolis & S. Laurence (Eds.), *The Conceptual Mind: New Directions in the Study of Concepts* (pp. 215–290). Cambridge, Massachusetts: MIT Press.
Evans, V., Bergen, B. K., & Zinken, J. (2007). *The cognitive linguistics enterprise: An overview*. 35.
Evans, V., & Green, M. (2006). *Cognitive Linguistics: An Introduction* (1 edizione). Mahwah; London; Edinburgh: Routledge.
Everett, C. (2013). *Linguistic Relativity: Evidence Across Languages and Cognitive Domains* (1 edizione). Berlin ; Boston: Mouton De Gruyter.
Evert, S. (2008). Corpora and collocations. In A. Lu¨deling & M. Kyto¨(Eds.), *Corpus Linguistics: An International Handbook* (pp. 1212–48). Berlin: de Gruyter.
Fauconnier, G. (1985). *Mental spaces: Aspects of meaning construction in natural language*. Cambridge University Press.
Fay, D., & Cutler, A. (1977). Malapropisms and the Structure of the Mental Lexicon. *Linguistic Inquiry*, *8*(3), 505–520.

Feng, Y., & Lapata, M. (2010). Visual Information in Semantic Representation. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 91–99. Retrieved from https://www.aclweb.org/anthology/N10-1011

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., … Welty, C. (2010). Building Watson: An Overview of the DeepQA Project. *AI Magazine*, *31*(3), 59–79. https://doi.org/10.1609/aimag.v31i3.2303

Fillmore, C. (1975). An alternative to checklist theories of meaning. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 123–131. Amsterdam: North Holland.

Fillmore, C. (1977). Scenes-and-frames semantics. In A. Zampoli (Ed.), *Linguistic Structures Processing* (pp. 55–82). Amsterdam: North Holland.

Fillmore, C. (1982). Frame semantics. In Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm* (pp. 111–137). Seoul: Hanshin Publishing.

Fillmore, C. (1985). Frames and the semantics of understanding. *Quaderni Di Semantica*, *6*, 222–254.

Fillmore, C., & Atkins, B. T. (1992). Toward a frame-based lexicon: The semantics of RISK and its neighbors. In A. Lehrer & E. F. Kittay (Eds.), *Frames, Fields and Contrasts* (pp. 75–102). Hillsdale, N.J: Lawrence Erlbaum Associates.

Fillmore, C., Kay, P., & O'Connor, M. C. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language*, *64*(3), 501–538.

Firth, J. R. (1957a). General linguistics and descriptive grammar. In *Papers in linguistics, 1934-1951* (pp. 216–228). Oxford University Press.

Firth, J. R. (1957b). Modes of meaning. In *Papers in linguistics, 1934-1951* (pp. 190–215). Oxford University Press.

Firth, J. R. (1957c). Personality and language in society. In *Papers in linguistics, 1934-1951* (pp. 177–189). Oxford University Press.

Firth, J. R. (1957d). The semantics of linguistic science. In *Papers in linguistics, 1934-1951* (pp. 139–147). Oxford University Press.

Firth, J. R. (1968a). A synonsis of linguistic theory, 1930-55. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth, 1952-59* (pp. 168–205). Indiana University Press.

Firth, J. R. (1968b). Descriptive linguistics and the study of meaning. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth, 1952-59* (pp. 96–113). Indiana University Press.

Firth, J. R. (1968c). Linguistic analysis and translation. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth, 1952-59* (pp. 74–83). Indiana University Press.

Firth, J. R. (1968d). Linguistic analysis as the study of meaning. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth, 1952-59* (pp. 12–26). Indiana University Press.

Firth, J. R. (1968e). Structural linguistics. In F. R. Palmer (Ed.), *Selected Papers of J. R. Firth, 1952-59* (pp. 35–52). Indiana University Press.

Fodor, J. A. (1975). *The Language of Thought* (1 edition). New York: Crowell.

Fodor, J. A. (1983). *Modularity of Mind*. Cambridge, Mass: Bradford Books.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong* (1 edition). Oxford : New York: Clarendon Press.

Fodor, J. D., Fodor, J. A., & Garrett, M. F. (1975). The Psychological Unreality of Semantic Representations. *Linguistic Inquiry*, *6*(4), 515–531.

Forster, K. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms: A collection of psycholinguistic studies* (pp. 257–287). Amsterdam ; New York: North-Holland Pub. Co.

Foxe, J., & Simpson, G. (2002). Flow of activation from V1 to frontal cortex in humans. *Experimental Brain Research*, *142*(1), 139–150. https://doi.org/10.1007/s00221-001-0906-7

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80–96. https://doi.org/10.1016/j.cogpsych.2014.08.002

Fromkin, V. A. (1985). Evidence in linguistics. In R. H. Robbins & V. A. Fromkin (Eds.), *Linguistics and linguistic evidence*. Newcastle upon Tyne: Grevatt & Grevat.

Fukkink, R. G. (2005). Deriving word meaning from written context: A process analysis. *Learning and Instruction*, *15*(1), 23–43. https://doi.org/10.1016/j.learninstruc.2004.12.002

Fukkink, R. G., Blok, H., & de Glopper, K. (2001). Deriving Word Meaning from Written Context: A Multicomponential Skill. *Language Learning*, *51*(3), 477–496. https://doi.org/10.1111/0023-8333.00162

Gallagher, S. (2008). Are Minimal Representations Still Representations? *International Journal of Philosophical Studies*, *16*(3), 351–369. https://doi.org/10.1080/09672550802113243

Gallese, V., & Lakoff, G. (2005). The Brain's concepts: The role of the Sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3), 455–479. https://doi.org/10.1080/02643290442000310

Galton, F. (1880). Statistics of Mental Imagery. *Mind*, *5*, 301–318.

Gambarara, D. (1999). Premessa. In *Semantica. Teorie, tendenze e problemi contemporanei* (pp. 11–24). Roma: Carocci.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Gargani, D. (2007). Saussure e la "massa amorfa." In A. Elia & M. De Palo (Eds.), *La lezione di Saussure. Saggi di epistemologia linguistica* (pp. 157–170). Roma: Carocci.

Garrett, M. F. (1975). The Analysis of Sentence Production. In G. Bower (Ed.), *Psychology of Learning and Motivation: Advances in research and theory* (Vol. 9, pp. 133–177). https://doi.org/10.1016/S0079-7421(08)60270-4

Garrett, M. F. (1980). Levels of processing in language production. In B. Butterworth (Ed.), *Language production* (Vol. 1). New York: Academic Press.

Garroni, E. (1998). L'indeterminatezza semantica: Una questione limitare. In F. Albano Leoni, D. Gambarara, S. Gensini, F. Lo Piparo, & R. Simone (Eds.), *Ai limiti del linguaggio* (pp. 49–78). Roma-Bari: Laterza.

Gasser, M., & Lee, C.-D. (1990). *Networks that learn phonology*. Indiana University: Computer Science Department.

Geckeler, H. (1971a). *Lexikalische Strukturen im Vergleich: Kontrastive Skizze zur Strukturierung des Wortfeldes "alt - jung- neu" im heutigen Italienisch, Spanisch und Französisch*. Max Niemeyer.

Geckeler, H. (1971b). *Zur Wortfelddiskussion: Untersuchungen zur Gliederung des Wort feldes "alt, jung, neu" im heutigen Französisch*. München: Wilhelm Fink Verlag.

Geckeler, H. (1979). *La semantica strutturale* (G. Klein, Trans.). Torino: Bollati Boringhieri.

Geeraerts, D. (1989). Prospects and problems of prototype theory. *Linguistics*, *27*(4), 587–612.
Geeraerts, D. (1993). Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics*, *4*(3), 223–272. https://doi.org/10.1515/cogl.1993.4.3.223
Geeraerts, D. (2006a). Introduction: A rough guide to Cognitive Linguistics. In D. Geeraerts (Ed.), *Cognitive Linguistics: Basic Readings* (pp. 1–28). Berlin: Mouton de Gruyter.
Geeraerts, D. (2006b). *Words and Other Wonders*. Berlin ; New York: Mouton de Gruyter.
Geeraerts, D. (2010). *Theories of Lexical Semantics*. Oxford University Press.
Geeraerts, D., Grondelaers, S., & Bakema, P. (1994). *The Structure of Lexical Variation: Meaning, Naming, and Context*. Walter de Gruyter.
Gensini, S. (2019). La semantica "integrata" di Tullio De Mauro. In T. De Mauro, *Il valore delle parole*. Roma: Ist. dell'Enciclopedia Italiana.
Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). https://doi.org/10.1017/CBO9780511620669.010
Gentner, D., & Goldin-Meadow, S. (2003a). *Language in Mind: Advances in the Study of Language and Thought*. The MIT Press.
Gentner, D., & Goldin-Meadow, S. (2003b). Whither Whorf. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 3–14). The MIT Press.
Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*(2), 249–268. https://doi.org/10.1017/S0305000904006786
Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 489–494. https://doi.org/10.1073/pnas.0509868103
Gilbert, C. D., & Sigman, M. (2007). Brain States: Top-Down Influences in Sensory Processing. *Neuron*, *54*(5), 677–696. https://doi.org/10.1016/j.neuron.2007.05.019
Gladstone, W. E. (William E. (1858). *Studies on Homer and the Homeric age*. Retrieved from http://archive.org/details/studiesonhomerho03glad
Gleitman, L. (2010). Economist Debates: Language: This House Believes that the Language We Speak Shapes How We Think. *Available at: Http://Www.Economist.Com/Debate/ Days/View/632 [Retrieved September 1, 2011]*.
Glenberg, A. M., & Mehta, S. (2008). Constraint on covariation: It's not meaning. *Rivista Di Linguistica*, *20*(1), 237–262.
Glenberg, A. M., & Robertson, D. A. (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language*, *43*, 379–401.
Glenberg, A., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes*, *28*(1), 1–26.
Goldberg, A. E. (1992). The Inherent Semantics of Argument Structure: The Case of the English Ditransitive Construction. *Cognitive Linguistics*, *3*(1), 37–74. https://doi.org/10.1515/cogl.1992.3.1.37
Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224. https://doi.org/10.1016/S1364-6613(03)00080-9

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.
Goldberg, Y., & Levy, O. (2014). word2vec Explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. *ArXiv:1402.3722 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1402.3722
Goldinger, S. D. (1998). Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review*, *105*(2), 251–279.
Goldin-Meadow, S. (2003). Thought before language: Do we think ergative? In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 493–522). Cambridge, MA, US: MIT Press.
Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*, 231–262. https://doi.org/10.1016/S0010-0277(97)00047-4
Grandy, R. E. (1992). Semantic Fields, Prototypes and the Lexicon. In A. Lehrer & E. F. Kittay (Eds.), *Frames, Fields and Contrasts* (pp. 103–122). Hillsdale, N.J: Lawrence Erlbaum Associates.
Greimas, A. J. (1966). *Sémantique structurale. Recherches de méthode*. Paris: Larousse.
Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. https://doi.org/10.1037/0033-295X.114.2.211
Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. *ArXiv:1803.11138 [Cs]*. Retrieved from http://arxiv.org/abs/1803.11138
Gumperz, J. J., & Levinson, S. C. (1996a). Introduction: Linguistic relativity re-examined. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 1–18). Cambridge University Press.
Gumperz, J. J., & Levinson, S. C. (Eds.). (1996b). *Rethinking linguistic relativity*. Cambridge University Press.
Günther, F., Dudschig, C., & Kaup, B. (2016). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *Quarterly Journal of Experimental Psychology (2006)*, *69*(4), 626–653. https://doi.org/10.1080/17470218.2015.1038280
Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *ArXiv:1605.09096 [Cs]*. Retrieved from http://arxiv.org/abs/1605.09096
Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6
Harris, C. L. (1992). Connectionism and Cognitive Linguistics. In N. Sharkey (Ed.), *Connectionist Natural Language Processing: Readings from Connection Science* (Softcover reprint of the original 1st ed. 1992 edizione, pp. 1–27). Springer Nature.
Harris, Z. S. (1954). Distributional Structure. *WORD*, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520
Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*(1), 10–20. https://doi.org/10.1037/h0032606
Hill, F., Reichart, R., & Korhonen, A. (2015). SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation. *Computational Linguistics*, *41*(4), 665–695. https://doi.org/10.1162/COLI_a_00237

Hill, J. H. (1988). Language, culture, and world view. In F. Newmeyer (Ed.), *Linguistics: The Cambridge survey*: *Vol. Language: the cultural context* (p. 23). Cambridge University Press.
Hill, J. H., & Mannheim, B. (1992). Language and World View. *Annual Review of Anthropology*, *2*, 381–406.
Hinton, G. E. (1981). Implementing semantic networks in parallel hardware. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 161–188). Hillsdale, NJ: Erlbaum.
Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed Representations. In D. E. Rumelhart, J. L. McClelland, & P. D. P. R. G. P. D. P. R. Group (Eds.), *Parallel Distributed Processing, Vol. 1: Foundations* (Reprint, pp. 77–109). Cambridge u.a: A Bradford Book.
Hjelmslev, L. (1961). *Prolegomena to a theory of language* (2nd Edition). University of Wisconsin Press.
Hjelmslev, L. (1981). *Saggi di linguistica generale*. Parma: Pratiche Editrice.
Hoffmann, T., & Trousdale, G. (2013a). Construction Grammar: Introduction. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar*. OUP USA.
Hoffmann, T., & Trousdale, G. (2013b). *The Oxford Handbook of Construction Grammar*. OUP USA.
Holmes, K. J., & Wolff, P. M. (2010). Simulation from schematics: Dorsal stream processing and the perception of implied motion. In R. Catrambone & S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (Vol. 32). Austin, Texas: Cognitive Science Society.
Hopman, E. W. M., Thompson, B., Austerweil, J. L., & Lupyan, G. (2018). Predictors of L2 word learning accuracy: A big data investigation. *Proceedings of the 40th Annual Conference of the Cognitive Science Society (CogSci 2018)*, 6.
Hulstijn, J.H. (1992). Retention of inferred and given word meaning: Experiments in incidental learning. In *Vocabulary and applied linguistics* (pp. 113–125). Basingstoke: Macmillan.
Hunn, E. S. (1977). *Tzeltal folk zoology: The classification of discontinuities in nature*. Academic Press.
Hurford, J. R. (1981). Malapropisms, Left-to-Right Listing, and Lexicalism. *Linguistic Inquiry*, *12*(3), 419–423.
Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813. https://doi.org/10.3758/BF03196544
Hymes, D. (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. University of Pennsylvania Press.
Iacobacci, I., Pilehvar, M. T., & Navigli, R. (2015). SensEmbed: Learning Sense Embeddings for Word and Relational Similarity. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 95–105. https://doi.org/10.3115/v1/P15-1010
Iacobacci, I., Pilehvar, M. T., & Navigli, R. (2016). Embeddings for Word Sense Disambiguation: An Evaluation Study. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 897–907. https://doi.org/10.18653/v1/P16-1085

Ipsen, G. (1924). Der alte Orient und die Indogermanen. In *Stand und Aufgaben der Sprachwissenschaft, Festschrift für Wilhelm Streitberg* (pp. 200–237). Heidelberg.
Jackendoff, R. (1983). *Semantics and Cognition*. MIT Press.
Jackendoff, R. (1987). The Status of Thematic Relations in Linguistic Theory. *Linguistic Inquiry*, *18*(3), 369–411. Retrieved from JSTOR.
Jackendoff, R. (1996). How Language Helps Us Think. *Pragmatics and Cognition*, *4*(1), 1–34. https://doi.org/10.1075/pc.4.1.03jac
Jackendoff, R. (2002). What's in the Lexicon? In S. Nooteboom, F. Weerman, & F. Wijnen (Eds.), *Storage and Computation in the Language Faculty* (pp. 23–58). https://doi.org/10.1007/978-94-010-0355-1_2
January, D., & Kako, E. (2007). *Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001) q*. 10.
Jarema, G., & Libben, G. (2007a). Introduction: Matters of Definition and Core Perspectives. In G. Jarema & G. Libben (Eds.), *The Mental Lexicon: Core Perspectives* (pp. 1–6). Amsterdam: Elsevier.
Jarema, G., & Libben, G. (Eds.). (2007b). *The Mental Lexicon: Core Perspectives*. Amsterdam: Elsevier.
Jenkins, J. J. (1954). Transitional organization: Association techniques. In C. E. Osgood & T. Sebeok (Eds.), *Psycholinguistics A Survey of Theory and Research Problems with A Survey of Psycholinguistic Research 1954-1964* (pp. 112–118). Bloomington: Indiana University Press.
Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning Vocabulary Through Reading. *American Educational Research Journ*, *21*(4), 21.
Jescheniak, J. D., & Levelt, W. J. M. (1994). Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843.
Ježek, E. (2005). *Lessico. Classi di parole, strutture, combinazioni*. Bologna: Il Mulino.
Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, *23*(4), 1214–1220. https://doi.org/10.3758/s13423-015-0980-7
Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, *17*(5), 662–672. https://doi.org/10.3758/PBR.17.5.662
Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press.
Jolles, A. (1934). Antike Bedeutungsfelder. *Beiträge Zur Geschichte Der Deutschen Sprache Und Literatur*, *58*, 97–109.
Jones, M. N., Gruenenfelder, T. M., & Recchia, G. (2018). In defense of spatial models of semantic representation. *New Ideas in Psychology*, *50*, 54–60. https://doi.org/10.1016/j.newideapsych.2017.08.001
Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552. https://doi.org/10.1016/j.jml.2006.07.003
Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford Handbook of Computational*

and Mathematical Psychology (pp. 232– 254). https://doi.org/10.1093/oxfordhb/9780199957996.001.0001
Joseph, J. E. (2000). *Limiting the Arbitrary: Linguistic naturalism and its opposites in Plato's Cratylus and modern theories of language*. https://doi.org/10.1075/sihols.96
Joseph, J. E. (2012). *Saussure*. Oxford ; New York: OUP Oxford.
Joseph, J. E. (2015). Iconicity in Saussure's Linguistic Work, and why it does not contradict the arbitrariness of the sign. *Historiographia Linguistica*, *42*(1), 85–105. https://doi.org/10.1075/hl.42.1.05jos
Joyce, T. (2005). *Constructing a Large-Scale Database of Japanese Word Associations*. *Glottometrics*(10), 82–98.
Jung, J., Na, L., & Akama, H. (2010). Network Analysis of Korean Word Associations. *Proceedings of the NAACL HLT 2010 First Workshop on Computational Neurolinguistics*, 27–35. Retrieved from http://dl.acm.org/citation.cfm?id=1866686.1866690
Katz, J. J. (1966). *The Philosophy of Language*. Harper & Row.
Katz, J. J. (1972). *Semantic theory*. New York: Harper & Row.
Katz, J. J., & Fodor, J. A. (1963). The Structure of a Semantic Theory. *Language*, *39*(2), 170. https://doi.org/10.2307/411200
Kay, P. (2004). NSM and the meaning of color words. *Theoretical Linguistics*, *29*(3). https://doi.org/10.1515/thli.29.3.237
Kay, P., & Kempton, W. (1984). What Is the Sapir-Whorf Hypothesis? *American Anthropologist*, *86*(1), 65–79. https://doi.org/10.1525/aa.1984.86.1.02a00050
Keil, F. C. (1979). *Semantic and Conceptual Development: An Ontological Perspective* (First Edition edition). Cambridge, Mass: Harvard University Press.
Keil, F. C., & Wilson, R. A. (Eds.). (1999). *The Mit Encyclopedia of Cognitive Science* (Revised edizione). Cambridge, Mass.: Bradford Books.
Kempen, G., & Huijbers, P. (1983). The lexlcalization process in sentence puoduction and naming: Indirect election of words. *Cognition*, *14*, 185–209.
Kenett, Y. N., Levi, E., Anaki, D., & Faust, M. (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1470–1489. https://doi.org/10.1037/xlm0000391
Kiela, D. (2017). *Deep embodiment: Grounding semantics in perceptual modalities* (No. UCAM-CL-TR-899). Retrieved from University of Cambridge, Computer Laboratory website: https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-899.html
Kiela, D., & Bottou, L. (2014). Learning Image Embeddings using Convolutional Neural Networks for Improved Multi-Modal Semantics. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 36–45. https://doi.org/10.3115/v1/D14-1005
Kiela, D., Bulat, L., & Clark, S. (2015). Grounding Semantics in Olfactory Perception. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 231–236. https://doi.org/10.3115/v1/P15-2038
Kiela, D., & Clark, S. (2014). A Systematic Study of Semantic Vector Space Model Parameters. *Proceedings of the 2nd Workshop on Continuous Vector Space Models and Their Compositionality (CVSC)*, 21–30. https://doi.org/10.3115/v1/W14-1503

Kiela, D., & Clark, S. (2015). Multi- and Cross-Modal Semantics Beyond Vision: Grounding in Auditory Perception. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2461–2470. https://doi.org/10.18653/v1/D15-1293

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. Aitken, R. Beiley, & N. Hamilton-Smith (Eds.), *The computer and literary studies*. Edinburgh: Edinburgh University Press.

Koerner, K. E. F. (1992). The Sapir-Whorf Hypothesis: A Preliminary History and a Bibliographical Essay. *Journal of Linguistic Anthropology*, *2*(2), 173–198.

Koerner, K. E. F. (2000). J. R. Firth and the Cours de linguistique générale: A Historiographical Sketch. In O. M. Tomić & M. Rodovanović (Eds.), *History and Perspectives of Language Study*: *Vol. Paper in honor of Ranko Bugarski* (pp. 115–132). Amsterdam & Philadelphia: John Benjamins Publishing Company.

Kohn, S. E., Wingfield, A., Menn, L., Goodglass, H., Gleason, J. B., & Hyde, M. (1987). Lexical retrieval: The tip-of-the-tongue phenomenon. *Applied Psycholinguistics*, *8*(3), 245–266. https://doi.org/10.1017/S0142716400000291

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent Processing in V1/V2 Contributes to Categorization of Natural Scenes. *Journal of Neuroscience*, *31*(7), 2488–2492. https://doi.org/10.1523/JNEUROSCI.3074-10.2011

Komatsu, L. K. (1992). Recent Views of Conceptual Structure. *Psychological Bulletin*, *112*(3), 500–526.

Kuhn, M. R., & Stahl, S. A. (1998). Teaching Children to Learn Word Meanings from Context: A Synthesis and Some Questions. *Journal of Literacy Research*, *30*(1), 119–138. https://doi.org/10.1080/10862969809547983

Labov, W. (1973). The boundaries of words and their meanings. In C.-J. Bailey & R. W. Shuy (Eds.), *New Ways of Analysing Variation in English* (pp. 340–371). Washington, DC: Georgetown University Press.

Labov, W. (1978). Denotational structure. In D. Farkas, W. M. Jacobsen, & K. W. Todrys (Eds.), *Papers from the Parasession on the Lexicon* (pp. 220–260). Chicago: Chicago Linguistic Society.

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.

Lakoff, G. (1990). The Invariance Hypothesis: Is Abstract Reason Based on Image-Schemas? *Cognitive Linguistics*, *I*, 39–44.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind & its Challenge to Western Thought*. New York, NY: Basic Books.

Lalumera, E. (2013). *Che cos'è il relativismo cognitivo*. Carocci.

Lancichinetti, A., Radicchi, F., Ramasco, J. J., & Fortunato, S. (2011). Finding Statistically Significant Communities in Networks. *PLoS ONE*, *6*(4), e18961. https://doi.org/10.1371/journal.pone.0018961

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, *104*(2), 211–240.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028
Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical prerequisites*. Stanford University Press.
Langacker, R. W. (1990). *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Walter de Gruyter.
Langacker, R. W. (1998). Conceptualization, Symbolization, and Grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 1–40). L. Erlbaum.
Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford ; New York: OUP USA.
Lapesa, G., & Evert, S. (2014). A Large Scale Evaluation of Distributional Semantic Models: Parameters, Interactions and Model Selection. *Transactions of the Association for Computational Linguistics*, *2*, 531–546. https://doi.org/10.1162/tacl_a_00201
Lapesa, G., & Evert, S. (2017). Large-scale evaluation of dependency-based DSMs: Are they worth the effort? *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 394–400. Retrieved from https://www.aclweb.org/anthology/E17-2063
Lapesa, G., Evert, S., & Schulte im Walde, S. (2014). Contrasting Syntagmatic and Paradigmatic Relations: Insights from Distributional Semantic Models. *Proceedings of the Third Joint Conference on Lexical and Computational Semantics (*SEM 2014)*, 160–170. https://doi.org/10.3115/v1/S14-1020
Laurence, S., & Margolis, E. (1999). Concepts and Cognitive Sciences. In E. Margolis & S. Laurence (Eds.), *Concepts: Core Readings* (pp. 3–81). MIT Press.
Lazaridou, A., Bruni, E., & Baroni, M. (2014). Is this a wampimuk? Cross-modal mapping between distributional semantics and the visual world. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1403–1414. https://doi.org/10.3115/v1/P14-1132
Lazaridou, A., Marelli, M., & Baroni, M. (2017). Multimodal Word Meaning Induction From Minimal Exposure to Natural Text. *Cognitive Science*, *41*, 677–705. https://doi.org/10.1111/cogs.12481
Lee, P. (1979). The Whorf Theory Complex: A critical reconstruction. In K. E. F. Koerner (Ed.), *Amsterdam Studies in the Theory and History of Linguistic Science* (p. 758). Retrieved from https://www.jstor.org/stable/1421810?origin=crossref
Lehrer, A. (1992). Names and Naming: Why We Need Fields and Frames. In A. Lehrer & E. F. Kittay (Eds.), *Frames, Fields and Contrasts* (pp. 123–142). Hillsdale, N.J: Lawrence Erlbaum Associates.
Lehrer, A., & Kittay, E. F. (Eds.). (1992). *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*. Hillsdale, N.J: Lawrence Erlbaum Associates.
Lemmens, M. (2015). Cognitive Semantics. In N. Riemer (Ed.), *The Routledge Handbook of Semantics* (pp. 90–105). Abingdon: Routledge.
Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Rivista Di Linguistica*, *20*(1), 1–31.

Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics*, *4*(1), 151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254
Lenci, A., Baroni, M., & Cazzolli, G. (2013). Una prima analisi delle norme semantiche BLIND. In L. Marotta, L. Meini, & M. Donati (Eds.), *Parlare senza vedere. Rappresentazioni semantiche nei non vedenti* (pp. 83–93). Pisa: ETS.
Lenci, A., Baroni, M., Cazzolli, G., & Marotta, G. (2013). BLIND: A set of semantic feature norms from the congenitally blind. *Behavior Research Methods*, *45*(4), 1218–1233. https://doi.org/10.3758/s13428-013-0323-4
Lenci, A., & Benotto, G. (2012). Identifying hypernyms in distributional semantic spaces. *Proceedings of the 1st Joint Conference on Lexical and Computational Semantics*, 75–79. Stroudsburg, PA: Assoc. Comput. Linguist.
Lenneberg, Eric H. (1953). Cognition in Ethnolinguistics. *Language*, *29*(4), 463. https://doi.org/10.2307/409956
Lenneberg, Eric Heinz, Chomsky, N., & Marx, O. (1967). *Biological Foundations of Language* (99th ed. edizione). New York: John Wiley & Sons Inc.
Lepschy, G. C. (1968). Introduzione. In L. Hjelmslev, *I fondamenti della teoria del linguaggio*. Torino: Giulio Einaudi Editore.
Lepschy, G. C. (1997). *La linguistica strutturale* (10 edizione). Torino: Einaudi.
Lepschy, G. C. (1998). Appunti per una rassegna saussuriana. In F. Albano Leoni, D. Gambarara, S. Gensini, F. Lo Piparo, & R. Simone (Eds.), *Ai limiti del linguaggio* (pp. 229–238). Roma-Bari: Laterza.
Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press.
Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The Time Course of Lexical Access in Speech Production: A Study of Picture Naming. *Psychological Review*, *98*(1), 122–142.
Levinson, S. C. (1996). Frames of reference and Molyneux's question: Crosslinguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language, speech, and communication. Language and space* (pp. 109–169). Cambridge, MA, US: MIT Press.
Levinson, S. C. (1997). From outer to inner space: Linguistic categories and non-linguistic thinking. In J. Nuyts & E. Pederson (Eds.), *Language, culture and cognition, Vol. 1. Language and conceptualization* (pp. 13–45). New York, NY: Cambridge University Press.
Levinson, S. C. (2003). *Space in language and cognition: Explorations in Cognitive Diversity*. Cambridge University Press.
Levy, O., & Goldberg, Y. (2014a). Dependency-Based Word Embeddings. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 302–308. https://doi.org/10.3115/v1/P14-2050
Levy, O., & Goldberg, Y. (2014b). Linguistic Regularities in Sparse and Explicit Word Representations. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, 171–180. https://doi.org/10.3115/v1/W14-1618
Levy, O., & Goldberg, Y. (2014c). *Neural Word Embedding as Implicit Matrix Factorization*. 9.
Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Transactions of the Association for Computational Linguistics*, *3*, 211–225. https://doi.org/10.1162/tacl_a_00134
Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*(3), 265–294. https://doi.org/10.1016/S0010-0277(02)00009-4

Lichtheim, L. (1885). On aphasia. *Brain*, *7*, 433–484.
Lin, E. L., & Murphy, G. L. (2001). Thematic Relations in Adults' Concepts. *Journal of Experimental Psychology: General*, *130*(1), 3–28.
Liu, H. (2011). Linguistic Complex Networks: A new approach to language exploration. *Grundlagenstud Kybern Geisteswiss*, *52*(4), 151–170.
Liuzza, M. T., Cimatti, F., & Borghi, A. M. (2010). *Lingue, corpo, pensiero: Le ricerche contemporanee*. Roma: Carocci.
Louwerse, M. M. (2007). Symbolic or embodied representations: A case for symbol interdependency. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis*.
Louwerse, M. M. (2018). Knowing the Meaning of a Word by the Linguistic and Perceptual Company It Keeps. *Topics in Cognitive Science*, *10*(3), 573–589. https://doi.org/10.1111/tops.12349
Louwerse, M. M., & Jeuniaux, P. (2008). Language comprehension is both embodied and symbolic. In M. de Vega, A. M. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 309–326). New York: Oxford University Press Oxford.
Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*(1), 96–104. https://doi.org/10.1016/j.cognition.2009.09.002
Love, N. (1986). The linguistic thought of J.R. Firth. *Stellenbosch Papers in Linguistics*, *15*, 31–60.
Lucas, M. (2000). Semantic priming without association: A meta-analytic review. *Psychonomic Bulletin & Review*, *7*(4), 618–630. https://doi.org/10.3758/BF03212999
Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. Wiley.
Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge University Press.
Lucy, J. A. (1996a). *Grammatical Categories and Cognition*. Cambridge: Cambridge University Press.
Lucy, J. A. (1996b). The scope of linguistic relativity: An analysis and review of empirical research. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 37–69). Cambridge University Press.
Lucy, J. A. (1997). The linguistics of color. In C. L. Hardin & L. Maffi (Eds.), *Color Categories in Language and Thought* (pp. 320–346). Cambridge: Cambridge University Press.
Lucy, J. A. (2005). Through the Window of Language: Assessing the Influence of Language Diversity on Thought. *Theoria*, *54*, 299–309.
Lucy, J. A. (2010). Language structure, lexical meaning, and cognition. In B. C. Malt & P. M. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 266–286). Oxford University Press.
Lucy, J. A. (2016). Recent Advances in the Study of Linguistic Relativity in Historical Context: A Critical Assessment. *Language Learning*, *66*(3), 487–515. https://doi.org/10.1111/lang.12195
Lucy, J. A., & Gaskin, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 257–283). Cambridge, UK ; New York: Cambridge University Press.

Lucy, J. A., & Gaskin, S. (2003). Interaction of Language Type and Referent Type in the Development of Nonverbal Classification Preferences. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind*. https://doi.org/10.7551/mitpress/4117.003.0023

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

Lupyan, G. (2006). Labels facilitate learning of novel categories. In A. Cangelosi, A. D. M. Smith, & K. R. Smith (Eds.), *Proceedings of the Sixth International Conference on the Evolution of Language* (pp. 190–197). https://doi.org/10.1142/9789812774262_0025

Lupyan, G. (2012a). Language Augmented Prediction. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00422

Lupyan, G. (2012b). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G. (2012c). What Do Words Do? Toward a Theory of Language-Augmented Thought. In *Psychology of Learning and Motivation* (Vol. 57, pp. 255–297). https://doi.org/10.1016/B978-0-12-394293-7.00007-8

Lupyan, G. (2016). The Centrality of Language in Human Cognition. *Language Learning*, *66*(3), 516–553. https://doi.org/10.1111/lang.12155

Lupyan, G., & Bergen, B. K. (2016). How Language Programs the Mind. *Topics in Cognitive Science*, *8*(2), 408–424. https://doi.org/10.1111/tops.12155

Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(2), 167–193. https://doi.org/10.1017/langcog.2014.21

Lupyan, G., & Christiansen, M. H. (2002). *Case, Word Order, and Language Learnability: Insights from Connectionist Modeling*. https://doi.org/10.4324/9781315782379-138

Lupyan, G., & Clark, A. (2015). Words and the World: Predictive Coding and the Language-Perception-Cognition Interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Lupyan, G., & Lewis, M. (2017). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience*, 1–19. https://doi.org/10.1080/23273798.2017.1404114

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology. General*, *141*(1), 170–186. https://doi.org/10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110

Luraghi, S., & Gaeta, L. (2003). Introduzione. In L. Gaeta & S. Luraghi (Eds.), *Introduzione alla linguistica cognitiva* (pp. 17–36). Roma: Carocci.

Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge University Press.

Lyons, J. (1977). *Semantics: Volume 1* (Revised edizione). Cambridge ; New York: Cambridge University Press.

Mack, A., & Rock, I. (1998). *Inattentional Blindness*. MIT Press.

Maier, M., & Abdel Rahman, R. (2018). Native Language Promotes Access to Visual Consciousness. *Psychological Science*, *29*(11), 1757–1772. https://doi.org/10.1177/0956797618782181

Majid, A. (2010). Words for parts of the body. In B. C. Malt & P. M. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 58–71). Oxford University Press.

Majid, A., Boster, J. S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*, *109*(2), 235–250. https://doi.org/10.1016/j.cognition.2008.08.009

Majid, A., Jordan, F., & Dunn, M. (2015). Semantic systems in closely related languages. *Language Sciences*, *49*, 1–18. https://doi.org/10.1016/j.langsci.2014.11.002

Majid, A., & Levinson, S. C. (2011). The Senses in Language and Culture. *The Senses and Society*, *6*(1), 5–18. https://doi.org/10.2752/174589311X12893982233551

Malhotra, N., Krosnick, J. A., & Haertel, E. (2007). *The Psychometric Properties of the GSS Wordsum Vocabulary Test* (GSS Methodological Report No. 11).

Malt, B. C. (1995). Category Coherence in Cross-Cultural Perspective. *Cognitive Psychology*, *29*(2), 85–148. https://doi.org/10.1006/cogp.1995.1013

Malt, B. C., Gennari, S., & Imai, M. (2010). Lexicalization Patterns and the World-to-Words Mapping. In B. C. Malt & P. M. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 29–57). Oxford University Press.

Malt, B. C., Gennari, S., Imai, M., Ameel, E., Saji, N., & Majid, A. (2015). Where are the concepts? What words can and can't reveal. In E. Margolis & S. Laurence (Eds.), *The Conceptual Mind: New Directions in the Study of Concepts* (pp. 291–237). Cambridge, Massachusetts: MIT Press.

Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language*, *49*(1), 20–42. https://doi.org/10.1016/S0749-596X(03)00021-4

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus Naming: Similarity and the Linguistic Categorization of Artifacts. *Journal of Memory and Language*, *40*(2), 230–262. https://doi.org/10.1006/jmla.1998.2593

Malt, B. C., & Wolff, P. M. (Eds.). (2010). *Words and the mind: How words capture human experience*. New York: Oxford University Press.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Manning, C. D., & Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Mass: Mit Pr.

Marconi, D. (1992). Semantica cognitiva. In M. Santambrogio (Ed.), *Introduzione alla filosofia analitica del linguaggio* (pp. 431–482). Roma-Bari: Laterza.

Margolis, E., & Laurence, S. (Eds.). (2015). *The Conceptual Mind: New Directions in the Study of Concepts*. Cambridge, Massachusetts: MIT Press.

Margolis, E., & Laurence, S. (2019). Concepts. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019). Retrieved from https://plato.stanford.edu/archives/sum2019/entries/concepts/

Markošová, M. (2008). Network model of human language. *Physica A: Statistical Mechanics and Its Applications*, *387*(2–3), 661–666. https://doi.org/10.1016/j.physa.2007.09.027

Marr, D. (1976). Early Processing of Visual Information. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *275*(942), 483–519. https://doi.org/10.1098/rstb.1976.0090
Marslen-Wilson, W. D. (1973). Linguistic Structure and Speech Shadowing at Very Short Latencies. *Nature*, *244*(5417), 522–523.
Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, *189*(4198), 226–228.
Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication*, *4*(1–3), 55–73. https://doi.org/10.1016/0167-6393(85)90036-6
Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71. https://doi.org/10.1016/0010-0277(80)90015-3
Marslen-Wilson, W. D., & Tyler, L. K. (1981). Central processes in speech understanding. *Philosophical Tronsu,Tions of the Royal Society of LoOOoll*, *B*(295), 317–332.
Mazzone, M. (1999). Significato e riferimento. Tra filosofia analitica e scienze cognitive. In *Semantica. Teorie, tendenze e problemi contemporanei*. Roma: Carocci.
McClelland, J. L., & Bybee, J. L. (2007). Gradience of Gradience: A reply to Jackendoff. *The Linguistic Review*, *24*(4). https://doi.org/10.1515/TLR.2007.019
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0
McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259
McCutchen, D., & Logan, B. (2011). Inside Incidental Word Learning: Children's Strategic Use of Morphological Information to Infer Word Meanings. *Reading Research Quarterly*, *46*(4), 334–349. https://doi.org/10.1002/RRQ.003
McDonald, S. A., & Ramscar, M. (2001). Testing the Distributional Hypothesis: The Influence of Context on Judgements of Semantic Similarity. *Proceedings of the Annual Meeting of the Cognitive Science Society, 23(23)*. Retrieved from https://escholarship.org/uc/item/6959p7b0
McNamara, T. P. (1992). Theories of Priming: I. Associative Distance and Lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1173–1190.
McRae, K., & Jones, M. N. (2013). Semantic Memory. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology*. https://doi.org/10.1093/oxfordhb/9780195376746.013.0014
Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, *44*(12), 1469–1481. https://doi.org/10.1037/0003-066X.44.12.1469
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207
Medin, D. L., & Wattenmaker, W. D. (1987). Category cohesiveness, theories, and cognitive archeology. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 25–62). New York, NY, US: Cambridge University Press.
Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex*, *48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv Preprint ArXiv:1301.3781*. Retrieved from http://arxiv.org/abs/1301.3781

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 3111–3119.

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*(1), 1–28. https://doi.org/10.1080/01690969108406936

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and Perception*. Belknap Press of Harvard University Press.

Mohammad, S. M., Dorr, B. J., Hirst, G., & Turney, P. D. (2013). Computing Lexical Contrast. *Computational Linguistics*, *39*(3), 555–590. https://doi.org/10.1162/COLI_a_00143

Monaghan, P., & Christiansen, M. H. (2008). Integration of multiple probabilistic cues in syntax acquisition. *Trends in Corpus Research: Finding Structure in Data*, 139–163.

Morris, C. W. (1938). *Foundations of the Theory of Signs* (Underling edizione). Chicago London: University of Chicago Press.

Morton, J. (1979). Facilitation in Word Recognition: Experiments Causing Change in the Logogen Model. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of Visible Language* (pp. 259–268). https://doi.org/10.1007/978-1-4684-0994-9_15

Morton, J. (1981). The Status of Information Processing Models of Language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *295*(1077), 387–396. https://doi.org/10.1098/rstb.1981.0147

Morton, J. (1982). Disintegrating the lexicon: An information processing approach. In E. Walker & M. F. Garrett (Eds.), *Perspectives on mental representation* (pp. 89–109). New Jersey: LEA.

Moss, H., & Older, L. (1996). *Birkbeck Word Association Norms*. Hove, UK: Psychology Press.

Most, S. B., Scholl, B. J., Clifford, E. R., & Simons, D. J. (2005). What You See Is What You Set: Sustained Inattentional Blindness and the Capture of Awareness. *Psychological Review*, *112*(1), 217–242. https://doi.org/10.1037/0033-295X.112.1.217

Murphy, G. L. (2002). *The Big Book of Concepts*. Bradford Book.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, *92*(3), 289–316. https://doi.org/10.1037/0033-295X.92.3.289

Murphy, M. L. (2003). *Semantic Relations and the Lexicon: Antonymy, Synonymy and other Paradigms*. Cambridge University Press.

Murphy, M. L. (2010). *Lexical Meaning*. Cambridge University Press.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1897). Learning Word Meanings From Context During Normal Reading. *American Educational Research Journal*, *24*(2), 34.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly*, *20*(2), 233–253. https://doi.org/10.2307/747758

Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, *17*(02), 357–374. https://doi.org/10.1017/S0305000900013817

Naigles, L. R., & Swensen, L. D. (2007). Syntactic Supports for Word Learning. In E. Hoff & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 212–231). https://doi.org/10.1002/9780470757833.ch11

Napps, S. E., & Fowler, C. A. (1987). Formal relationships among words and the organization of the mental lexicon. *Journal of Psycholinguistic Research*, *16*(3), 257–272. https://doi.org/10.1007/BF01067546

Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys*, *41*(2), 1–69. https://doi.org/10.1145/1459352.1459355

Navigli, R. (2012). A Quick Tour of Word Sense Disambiguation, Induction and Related Approaches. In M. Bieliková, G. Friedrich, G. Gottlob, S. Katzenbeisser, & G. Turán (Eds.), *SOFSEM 2012: Theory and Practice of Computer Science* (Vol. 7147, pp. 115–129). https://doi.org/10.1007/978-3-642-27660-6_10

Navigli, R., & Martelli, F. (2019). An overview of word and sense similarity. *Natural Language Engineering*, 1–22. https://doi.org/10.1017/S1351324919000305

Neisser, Ulric, & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, *7*(4), 480–494. https://doi.org/10.1016/0010-0285(75)90019-5

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. https://doi.org/10.3758/BF03195588

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L. (1992). Processing Implicit and Explicit Representations. *Psychological Review*, *99*(2), 322–348.

Nematzadeh, A., Meylan, S. C., & Griffiths, T. L. (2017). Evaluating Vector-Space Models of Word Representation, or, The Unreasonable Effectiveness of Counting Words Near Other Words. *CogSci*.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, *45*(2), 167–256. https://doi.org/10.1137/S003614450342480

Ogarkova, A. (2016). Translatability of Emotions. In *Emotion Measurement* (pp. 575–599). https://doi.org/10.1016/B978-0-08-100508-8.00023-0

Ogden, C. K., & Richards, I. A. (1923). *Meaning Of Meaning* (First edition). San Diego: Mariner Books.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, Mass: Bradford Books.

Ouyang, L., Boroditsky, L., & Frank, M. C. (2017). Semantic Coherence Facilitates Distributional Learning. *Cognitive Science*, *41*, 855–884. https://doi.org/10.1111/cogs.12360

Özgen, E., & Davies, I. R. L. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, *131*(4), 477–493. https://doi.org/10.1037/0096-3445.131.4.477

Padó, S., & Lapata, M. (2007). Dependency-Based Construction of Semantic Space Models. *Computational Linguistics*, *33*(2), 161–199. https://doi.org/10.1162/coli.2007.33.2.161

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford University Press.

Pal, A. R., & Saha, D. (2015). Word sense disambiguation: A survey. *International Journal of Control Theory and Computer Modeling*, *5*(3), 1–16. https://doi.org/10.5121/ijctcm.2015.5301

Palmer, F. R. (1976). *Semantics: A New Outline*. Cambridge University Press.

Paradis, C. (2003). Is the notion of *linguistic competence* relevant inCognitive Linguistics? *Annual Review of Cognitive Linguistics*, *1*, 207–231. https://doi.org/10.1075/arcl.1.12par

Paradis, C. (2005). Ontologies and Construals in Lexical Semantics. *Axiomathes*, *15*(4), 541–573. https://doi.org/10.1007/s10516-004-7680-7
Paradis, C. (2012). Lexical Semantics. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (p. wbeal0695). https://doi.org/10.1002/9781405198431.wbeal0695
Peacocke, C. (1992). *A Study of Concepts* (New edition edition). Cambridge, Mass.: A Bradford Book.
Pederson, E. (1995). Language as Context, Language as Means: Spatial Cognition and Habitual Language Use. *Cognitive Linguistics*, *6*(1), 33–62. https://doi.org/10.1515/cogl.1995.6.1.33
Pederson, E., Danziger, E., Wilkins, D., Levinson, S. C., Kita, S., & Senft, G. (1998). Semantic Typology and Spatial Conceptualization. *Language*, *74*(3), 557. https://doi.org/10.2307/417793
Peirsman, Y. (2008). Word space models of semantic similarity and relatedness. *Proceedings of the 13th ESSLLI Student Session*, 143–152.
Peirsman, Y., Kris, H., & Speelman, D. (2007, January 1). *Finding semantically related words in Dutch. Co-occurrences versus syntactic contexts.* 9–16.
Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162
Pereira, F., Gershman, S., Ritter, S., & Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology*, *33*(3–4), 175–190. https://doi.org/10.1080/02643294.2016.1176907
Perry, L. K., & Lupyan, G. (2013). What the online manipulation of linguistic activity can tell us about language and thought. *Frontiers in Behavioral Neuroscience*, *7*. https://doi.org/10.3389/fnbeh.2013.00122
Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, *135*, 66–72. https://doi.org/10.1016/j.bandl.2014.05.005
Phillips, W., & Boroditsky, L. (2003). Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Concepts. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Societyi*. Mahwah, NY: Lawrence Erlbaum Associates.
Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, *21*(5), 1112–1130. https://doi.org/10.3758/s13423-014-0585-6
Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137–158). John Benjamins Publishing.
Pierrehumbert, J. B. (2003). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech*, *46*(2–3), 115–154. https://doi.org/10.1177/00238309030460020501
Pinker, S. (1984). *Language learnability and language development*. Harvard University Press.
Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York, NY: Harper Perennial Modern Classics.
Pinker, S. (2007). *The Stuff of Thought: Language as a Window Into Human Nature*. Viking.

Plaut, D. C. (1997). Structure and Function in the Lexical System: Insights from Distributed Models of Word Reading and Lexical Decision. *Language and Cognitive Processes*, *12*(5–6), 765–806. https://doi.org/10.1080/016909697386682

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-Regular Domains. *Psychological Review*, *103*, 56–115.

Porzig, W. (1934). Wesenhafte Bedeutungsbeziehunge. *Beiträge Zur Deutschen Sprache Und Literatur*, *58*, 70–79.

Postman, L., & Keppel, G. (Eds.). (1970). *Norms of Word Association*. Academic Press.

Prampolini, M. (1998). Il concetto di vaghezza. In F. Albano Leoni, D. Gambarara, S. Gensini, F. Lo Piparo, & R. Simone (Eds.), *Ai limiti del linguaggio* (pp. 97–110). Roma-Bari: Laterza.

Prieto, L. J. (1964). *Principes de Nooologie. Fondements de la théorie fonctionnelle du si- gnifié*. Paris: Mouton, The Hague.

Prinz, J. J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. MIT Press.

Pullum, G. K. (1989). The great Eskimo vocabulary hoax. *Natural Language & Linguistic Theory*, *7*(2), 275–281.

Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, Mass.: MIT Press.

Pylyshyn, Z. W. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(03). https://doi.org/10.1017/S0140525X99002022

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, *12*(5), 410–430. https://doi.org/10.1002/bs.3830120511

Quillian, M. R. (1969). The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM*, *12*(8), 459–476. https://doi.org/10.1145/363196.363214

Rambelli, G., Chersoni, E., Blache, P., Huang, C.-R., & Lenci, A. (2019). Distributional Semantics Meets Construction Grammar. Towards a Unified Usage-Based Model of Grammar and Meaning. *First International Workshop on Designing Meaning Representations*, 12. Florence, Italy.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *The Quarterly Journal of Experimental Psychology Section A*, *55*(4), 1339–1362. https://doi.org/10.1080/02724980244000099

Reali, F., Christiansen, M. H., & Monaghan, P. (2003). Phonological and Distributional Cues in Syntax Acquisition: Scaling up the Connectionist Approach to Multiple-Cue Integration. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25). Retrieved from https://escholarship.org/uc/item/85355384

Redington, M., Chater, N., & Finch, S. (1998). Distributional Information: A Powerful Cue for Acquiring Syntactic Categories. *Cognitive Science*, *22*(4), 425–469. https://doi.org/10.1207/s15516709cog2204_2

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439–446. https://doi.org/10.1016/j.tics.2009.07.001

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences*, *102*(23), 8386–8391. https://doi.org/10.1073/pnas.0503281102

Regier, T., Kay, P., Gilbert, A. L., & Ivry, R. B. (2010). Language and thought: Which side are you on, anyway? In B. C. Malt & P. M. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 165–182). Oxford University Press.

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, *27*(1), 611–647. https://doi.org/10.1146/annurev.neuro.26.041002.131039

Riemer, N. (2006). Reductive Paraphrase and Meaning: A Critique of Wierzbickian Semantics. *Linguistics and Philosophy*, *29*(3), 347–379.

Riemer, N. (2013). Conceptualist semantics: Explanatory power, scope and uniqueness. *Language Sciences*, *35*, 1–19. https://doi.org/10.1016/j.langsci.2012.09.003

Riemer, N. (2015). Meaning, conceptualization and expression. In N. Riemer (Ed.), *The Routledge Handbook of Semantics* (pp. 30–47). Abingdon: Routledge.

Riordan, B., & Jones, M. N. (2011). Redundancy in Perceptual and Linguistic Experience: Comparing Feature-Based and Distributional Models of Semantic Representation. *Topics in Cognitive Science*, *3*(2), 303–345. https://doi.org/10.1111/j.1756-8765.2010.01111.x

Roberson, D. (2005). Color Categories Are Culturally Diverse in Cognition as Well as in Language. *Cross-Cultural Research*, *39*(1), 56–71. https://doi.org/10.1177/1069397104267890

Rodda, M. A., Senaldi, M. S. G., & Lenci, A. (2016). Panta rei: Tracking Semantic Change with Distributional Semantics in Ancient Greek. In A. Corazza, S. Montemagni, & G. Semeraro (Eds.), *Proceedings of the Third Italian Conference on Computational Linguistics CLiC-it 2016* (pp. 258–262). https://doi.org/10.4000/books.aaccademia.1835

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47*, 59–87.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. MIT Press.

Rosch, E. (1973). Natural Categories. *Cognitive Psychology*, *4*, 328–350.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Rumelhart, D. E. (1979). Some problems with the notion of literal meanings. In *Metaphor and thought* (Vol. 2, pp. 71-82.).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Internal Representations by Error Propagation. In D. E. Rumelhart, J. L. McClelland, & P. D. P. R. G. P. D. P. R. Group (Eds.), *Parallel Distributed Processing, Vol. 1: Foundations* (Reprint, pp. 318–362). Cambridge u.a: A Bradford Book.

Rumelhart, D. E., & McClelland, J. L. (1986). On Learning the Past Tense of English Verbs. In J. L. McClelland, D. E. Rumelhart, & P. R. Group (Eds.), *Parallel Distributed Processing, Vol. 2: Psychological and Biological Models* (pp. 216–271). Cambridge: A Bradford Book.

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In *Mechanisms of language acquisition* (pp. 195–248). Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D. E., McClelland, J. L., & Group, P. D. P. R. G. P. D. P. R. (1986). *Parallel Distributed Processing, Vol. 1: Foundations* (Reprint). Cambridge u.a: A Bradford Book.
Saeed, J. I. (1997). *Semantics*. John Wiley & Sons.
Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language*, *47*(1), 172–196. https://doi.org/10.1006/jmla.2001.2839
Saffran, J. R. (2003). Statistical Language Learning: Mechanisms and Constraints. *Current Directions in Psychological Science*, *12*(4), 110–114. https://doi.org/10.1111/1467-8721.01243
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926
Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032
Sahlgren, M. (2006). *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector space*. Stockholm University, Stockholm.
Sahlgren, M. (2008). The distributional hypothesis. *Rivista Di Linguistica*, *20*(1), 33–53.
Sandra, D. (1994). The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. *Language and Cognitive Processes*, *9*(3), 227–269. https://doi.org/10.1080/01690969408402119
Santus, E., Lenci, A., Lu, Q., & Schulte im Walde, S. (2014). Chasing Hypernyms in Vector Spaces with Entropy. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Volume 2: Short Papers*, 38–42. https://doi.org/10.3115/v1/E14-4008
Santus, E., Lu, Q., Lenci, A., & Huang, C.-R. (2014). Taking Antonymy Mask off in Vector Space. *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing*, 135–144. Retrieved from https://www.aclweb.org/anthology/Y14-1018
Santus, E., Yung, F., Lenci, A., & Huang, C.-R. (2015). EVALution 1.0: An Evolving Semantic Dataset for Training and Evaluation of Distributional Semantic Models. *Proceedings of the 4th Workshop on Linked Data in Linguistics: Resources and Applications*, 64–69. https://doi.org/10.18653/v1/W15-4208
Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. Harcourt, Brace.
Sapir, E. (1924). The Grammarian and his Language. *The American Mercury*, *1*, 149–155.
Sapir, E. (1929). The Status of Linguistics as a Science. *Language*, *5*(4), 207. https://doi.org/10.2307/409588
Saunders, B. (2000). Revisiting Basic Color Terms. *Journal of the Royal Anthropological Institute*, *6*(1), 81–99. https://doi.org/10.1111/1467-9655.00005
Saussure, F. de. (1967). *Corso di linguistica generale* (T. De Mauro, Ed.). Roma-Bari: Laterza.
Saussure, F. de. (1983). *Course in General Linguistics* (R. Harris, Ed.). Bloomsbury Academic.
Saussure, F. de. (2005). *Scritti inediti di linguistica generale* (T. De Mauro, Trans.). Roma; Bari: Laterza.
Saussure, F. de. (2006). *Nouveaux regards sur Saussure: Mélanges offerts à René Amacker* (R. Engler, R. Simone, C. Touratier, A. Rousseau, & Collectif, Eds.). Genève: DROZ.
Schriefers, H., & Vigliocco, G. (2001). Speech Production, Psychology of. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 255–258). https://doi.org/10.1016/B978-0-08-097086-8.52022-4

Schütze, H. (1997). *Ambiguity Resolution in Language Learning: Computational and Cognitive Models*. Stanford, Calif: Center for the Study of Language and Inf.
Schütze, H. (1998). Automatic Word Sense Discrimination. *Computational Linguistics*, *24*(1), 97–123.
Ščur, G. S. (1974). *Teorii polja v lingvistike*. Mosca: Nauka.
Searle, J. R. (1980). Mind, Brains and Programs. *Behavioral and Brain Sciences*, *3*(3), 417–457.
Seidenberg, M. S. (1997). Language Acquisition and Use: Learning and Applying Probabilistic Constraints. *Science*, *275*, 1599–1603.
Seidenberg, M. S., & McClelland, J. L. (1989). A Distributed, Developmental Model of Word Recognition and Naming. *Psychological Review*, *96*(4), 532–568.
Shallice, T. (1979). Case study approach in neuropsychological research. *Journal of Clinical Neuropsychology*, *1*(3), 183–211. https://doi.org/10.1080/01688637908414450
Shapiro, L. (2008). Symbolism, embodied cognition, and the broader debate. In M. de Vega, A. Glenberg, & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition*. https://doi.org/10.1093/acprof:oso/9780199217274.001.0001
Shwartz, V., Santus, E., & Schlechtweg, D. (2017). Hypernyms under Siege: Linguistically-motivated Artillery for Hypernymy Detection. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, 65–75. Retrieved from https://www.aclweb.org/anthology/E17-1007
Simone, R. (1992). Il corpo del linguaggio. Il paradigma dell'arbitrarietà e il paradigma della sostanza. In R. Simone (Ed.), *Il sogno di Saussure* (pp. 37–59). Roma-Bari: Laterza.
Sinclair, J. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.
Slobin, D. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.
Slobin, D. (2003). Language and Thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in Mind: Advances in the Study of Language and Thought* (pp. 175–191). MIT Press.
Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Harvard University Press.
Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*, 1–74.
Solomon, null, Medin, null, & Lynch, null. (1999). Concepts do more than categorize. *Trends in Cognitive Sciences*, *3*(3), 99–105.
Stemberger, J. P. (1985). *The lexicon in a model of language production*. New York: Garland Publishing.
Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, *29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3
Swanborn, M. S. L., & De Glopper, K. (1999). Incidental Word Learning While Reading: A Meta-Analysis. *Review OfEducational Resea*, *69*(3), 25.
Szalay, L. B., & Deese, J. (1978). *Subjective meaning and culture: An assessment through word associations*. L. Erlbaum Associates.
Talmy, L. (1983). How Language Structures Space. In H. L. Pick & L. P. Acredolo (Eds.), *Spatial Orientation: Theory, Research, and Application* (pp. 225–282). https://doi.org/10.1007/978-1-4615-9325-6_11
Talmy, L. (2000). *Toward a Cognitive Semantics*. MIT Press.

Taylor, J. R. (1999). Cognitive semantics and structural semantics. In A. Blank & P. Koch (Eds.), *Historical Semantics and Cognition* (pp. 17–48). Walter de Gruyter.
Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences*, *113*(16), 4530–4535. https://doi.org/10.1073/pnas.1523631113
Thompson, B., Roberts, S., Roberts, S., & Lupyan, G. (2018). Quantifying Semantic Alignment Across Languages. *Proceedings of the 40th Annual Conference of the Cognitive Science Society (CogSci 2018)*, 2551–2556.
Thompson, G. W., & Kello, C. T. (2014). Walking across Wikipedia: A scale-free network model of semantic memory retrieval. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00086
Tomasello, M. (1998a). Introduction: A Cognitive-Functional Perspective on Language Structure. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. VII–XXIII). L. Erlbaum.
Tomasello, M. (1998b). *The new psychology of language: Cognitive and functional approaches to language structure*. L. Erlbaum.
Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–B34. https://doi.org/10.1016/j.cognition.2005.01.006
Treisman, A. M. (1961). *Attention and speech*. Oxford University, Unpuslished doctoral dissertation.
Trier, J. (1931). *Der deutsche Wortschatz im Sinnbezirk des Verstandes; die Geschichte eines Sprachlichen feldes*. Heidelberg.
Turing, A. M. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *s2-42*(1), 230–265. https://doi.org/10.1112/plms/s2-42.1.230
Turney, P. D. (2008). A Uniform Approach to Analogies, Synonyms, Antonyms, and Associations. *ArXiv:0809.0124 [Cs]*. Retrieved from http://arxiv.org/abs/0809.0124
Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188. https://doi.org/10.1613/jair.2934
Utsumi, A. (2015). A Complex Network Approach to Distributional Semantic Models. *PLOS ONE*, *10*(8), e0136277. https://doi.org/10.1371/journal.pone.0136277
Van de Cruys, T. (2008). A Comparison of Bag of Words and Syntax-based Approaches for Word Categorization. *Proceedings of the Lexical Semantics Workshop: Bridging the Gap between Semantic Theory and Computational Simulations*, 47–54. Amsterdam: Amsterdam University.
van Staden, M., & Majid, A. (2006). Body colouring task. *Language Sciences*, *28*(2–3), 158–161. https://doi.org/10.1016/j.langsci.2005.11.004
Varela, F. J., Thompson, E. T., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience* (Revised ed. edition). Cambridge, Mass.: The MIT Press.
Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(02), 219–247. https://doi.org/10.1515/LANGCOG.2009.011
Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of*

Experimental Psychology. General, *134*(4), 501–520. https://doi.org/10.1037/0096-3445.134.4.501
Violi, P. (1997). *Significato ed esperienza*. Milano: Bompiani.
Ward, E. J., & Scholl, B. J. (2015). Inattentional blindness reflects limitations on perception, not memory: Evidence from repeated failures of awareness. *Psychonomic Bulletin & Review*, *22*(3), 722–727. https://doi.org/10.3758/s13423-014-0745-8
Warrington, E. K. (1975). The Selective Impairment of Semantic Memory. *Quarterly Journal of Experimental Psychology*, *27*(4), 635–657. https://doi.org/10.1080/14640747508400525
Weisgerber, L. (1926). Das Problem der inneren Sprachform. *GRM*, *XV*, 241–256.
Weisgerber, L. (1950). *Vom Weltbild der deutschen Sprache*. Dusseldorf: Schwann.
Weisgerber, L. (1954). Die Sprachfelder in der geistein Erschliessung der Welt. In *Festschrift Trier*. Hain: Mesenheim.
Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Discussions. *Infant Behavioral Development*, *7*, 49–63. https://doi.org/10.1016/s0163-6383(84)80022-3
Wernicke, C. (1874). *Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis*. Retrieved from http://archive.org/details/deraphasischesy00werngoog
Whitaker, H. A. (2006). Words in the Mind, Words in the Brain: Preface to inaugural issue of The Mental Lexicon. *The Mental Lexicon*, *1*(1), 3–5. https://doi.org/10.1075/ml.1.1.02whi
Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). MIT Press.
Wierzbicka, A. (1972). *Semantic primitives*. Athenäum-Verl.
Wierzbicka, A. (1985). *Lexicography and conceptual analysis*. Karoma.
Wierzbicka, A. (1996). *Semantics: Primes and Universals*. Oxford University Press, UK.
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104
Wittgenstein, L. (1958). *Philosophical investigations* (G. E. M. Anscombe, Ed.). Oxford: Basil Blackwell.
Wolff, P. M., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/wcs.104
Wolff, P. M., & Malt, B. C. (2010). The language-thought interface. An introduction. In P. M. Wolff & B. C. Malt (Eds.), *Words and the mind: How words capture human experience* (pp. 3–15). New York: Oxford University Press.
Wolff, P. M., Medin, D. L., & Pankratz, C. (1999). Evolution and devolution of folkbiological knowledge. *Cognition*, *73*(2), 177–204.
Wysocki, K., & Jenkins, J. R. (1987). Deriving Word Meanings through Morphological Generalization. *Reading Research Quarterly*, *22*(1), 66. https://doi.org/10.2307/747721
Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., … Bhattacharya, T. (2016). On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences*, *113*(7), 1766–1771. https://doi.org/10.1073/pnas.1520752113
Zlatev, J., & Blomberg, J. (2015). Language may indeed influence thought. *Frontiers in Psychology*, *6*, 1–10. https://doi.org/10.3389/fpsyg.2015.01631