# Exploring the Dissociated Nucleus Phenomenon in Semantic Role Labeling

Tommaso Bonomo[1,*], Simone Conia[1] and Roberto Navigli[1]

[1] *Sapienza University of Rome, Dipartimento di Ingegneria Informatica, Automatica e Gestionale (DIAG), Via Ariosto 25, Rome, Italy*

**Abstract**

Dependency-based Semantic Role Labeling (SRL) is bound to dependency parsing, as the arguments of a predicate are identified through the token that heads the dependency relation subtree of the argument span. However, dependency-based SRL corpora are susceptible to the *dissociated nucleus* problem: when a subclause's semantic and structural cores are two separate words, the dependency tree chooses the structural token as the head of the subtree, coercing the SRL annotation into making the same choice. This leads to undesirable consequences: when directly using the output of a dependency-based SRL method in downstream tasks it is useful to work with the token representing the semantic core of a subclause, not the structural core. In this paper, we carry out a linguistically-driven investigation on the *dissociated nucleus* problem in dependency-based SRL and propose a novel algorithm that aligns predicate-argument structures to the syntactic structures from Universal Dependencies to select the semantic core of an argument. Our analysis shows that *dissociated nuclei* appear more often than one might expect, and that our novel algorithm greatly increases the richness of the semantic information in dependency-based SRL. We release the software to reproduce our experiments at https://github.com/SapienzaNLP/semdepalign.

**Keywords**

Semantic Role Labeling, Dependency Parsing

## 1. Introduction

Within the field of Natural Language Processing, Semantic Role Labeling [1, SRL] is aimed at recognizing the semantic information conveyed by a sentence, more specifically identifying *who* did *what* to *whom*, *when*, *where* and *how* [2]. Over the years, SRL has split into two main annotation formalisms, namely, span-based and dependency-based. The key difference between the two lies in how they identify the roles of a predicate: span-based SRL directly extracts a span of the input text as the argument of a predicate, whilst dependency-based SRL identifies the word that heads the syntactic dependency relation subtree corresponding to the argument span as the argument. Using dependency-based SRL can be beneficial in real-world settings, as i) dependency-based SRL parsers have achieved better results on standard benchmarks, and ii) the identified token can be directly utilized in several downstream tasks, including Coreference Resolution [3], Opinion Role Labeling [4, 5], Argument Mining [6, 7], and Concept Map Mining [8], among others.

However, the use of role tokens in the above tasks requires them to carry the "semantic meaning" of the role. This requirement is often not fulfilled when examining both the output of state-of-the-art dependency-based SRL systems and the corpora they were trained on, such as CoNLL-2009 [9]. In these annotations, it is not uncommon to have an adpositional clause serving as the head word of a semantic role, even though adpositions do not represent the semantic core of that role. In linguistics, this phenomenon is referred to as an instance of *dissociated nucleus* [10, ch. 23]. Although this term encompasses many different syntactical constructions, here we focus on adpositional clauses present in the CoNLL-2009 dataset, across all of its languages.

In this paper, we carry out a concise, linguistically-driven investigation on dissociated nuclei in dependency-based SRL, uncovering the extent of this problem and how it affects the semantic aspect of this task. In addtion, we introduce SemDepAlign, a simple yet effective algorithm capable of mitigating this phenomenon significantly by aligning predicate-argument structures in SRL with syntactic parses from the Universal Dependencies project, which addresses the dissociated nucleus phenomenon directly in the dependency structures. Applying SemDepAlign to CoNLL-2009 results in a substantial increase in the semantic variety of role tokens, measured through a set of proxy metrics. Finally, we provide a glimpse at how addressing dissociated nuclei simplifies the alignment between Semantic Role Labeling and Semantic Parsing, specifically with Abstract Meaning Representation [11, AMR]. We release SemDepAlign and Aligned-CoNLL09 – the result of applying SemDepAlign to CoNLL-2009 – in the hope that our work can encourage a deeper focus on semantics in SRL and foster future integration of this task into downstream applications.

*CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Dec 04 — 06, 2024, Pisa, Italy*
*Corresponding author.
bonomo@diag.uniroma1.it (T. Bonomo); conia@diag.uniroma1.it (S. Conia); navigli@diag.uniroma1.it (R. Navigli)
0000-0002-6056-9982 (T. Bonomo); 0000-0002-6238-7816 (S. Conia); 0000-0003-3831-9706 (R. Navigli)

## 2. SRL and Dependency Parsing

Both SRL and Dependency Parsing investigate how words in the same sentence relate to each other, respectively in a semantic or syntactic sense. The Conference on Computational Natural Language Learning (CoNLL) organized several Shared Tasks regarding both tasks, culminating in the CoNLL-2008 Shared Task [12] that asked participants to identify both types of relation within an English-only corpus. This task can be seen as the first occurrence of dependency-based SRL, as it explicitly ties the SRL annotations to the dependency relation tree of the sentence. The authors of the Shared Task implemented their own constituency-to-dependency parser to obtain the syntactic dependency relation trees, which are vulnerable by construction to the dissociated nucleus phenomenon.

The dependency relation annotation scheme adopted in both CoNLL-2008 and its multilingual successor CoNLL-2009 [9] impacts the output of dependency-based SRL systems trained on these training sets. If one inspects either a training sample of CoNLL-2009 or an output of a system trained on it, one can expect to encounter the dissociated nucleus phenomenon [10, ch. 23]. For example, the training sample "That is a service to the nation" presents a dissociated nucleus: the structural and semantic functions of the subclause "to the nation" are fulfilled by two separate tokens, 'to' and 'nation', respectively. The annotation provided within CoNLL-2009 identifies the syntactic core 'to' with the argument A2 for the nominal predicate 'service' because it is the head of the original dependency relation subtree corresponding to the argument span. Consequently, many tokens annotated as arguments are simple adpositions of little semantic significance. This significant detail impacts downstream tasks that use SRL outputs as input: if we wanted to extract relations or perform disambiguation on the example above, we would have much more interest in focusing on the word 'nation' than the adposition 'to'.

A way to quantify this phenomenon is to look at the frequency of part-of-speech (POS) tags of role tokens in the corpus. We are interested in the POS label of "Preposition or subordinating conjunction", which is the second-most frequent tag with 76,821 role tokens out of a total of 475,069, or ~17% of all the role tokens. Table 5 in the Appendix provides a complete breakdown over all POS classes in the English-split of CoNLL-2009.

We argue that both the training corpora and dependency-based SRL systems should identify the semantic core of an argument span as the head of the argument. In Appendix A we provide further examples of this phenomenon in non-English partitions of CoNLL-2009.

**Algorithm 1:** SemDepAlign

**input:** the role node *role_node*; the root node of the UD dep-tree *root_ud*.
**output:** the head node of the role in the UD dep-tree.

*role_tokens* ← $get\_tokens$(*role_node*)
*ud_role_subtree* ← *root_ud*
*min_nodes* ← $SymDiff$($get\_tokens$(*root_ud*), *role_tokens*)
**for** *node* ← *BFS(root_ud)***:**
    *subtree_tokens* ← $get\_tokens$(node)
    *extra_nodes* ← $SymDiff$(*subtree_tokens*, *role_tokens*)
    *min_nodes* ← min(*min_nodes*, *extra_nodes*)
**return** *ud_role_subtree*

## 3. Re-associating Dissociated Nuclei

Having established that the current annotations in CoNLL-2009 are susceptible to the dissociated nucleus phenomenon, we aim to mitigate this issue by introducing a subtree alignment algorithm that leverages the characteristics of Universal Dependencies [13, 14, UD] to collapse arguments that have been placed on structural tokens with their corresponding semantic tokens. UD explicitly addresses the dissociated nucleus issue by extending the definition of a nominal to encompass the entire nominal extended projection, following the linguistic theory proposed by Grimshaw [15]. The nominal head is used as the referential core and the adposition is treated as a functional marker [14, Section 3.1.1]. When constructing the dependency tree structures, UD guidelines [14, Section 2.1.1] indicate that the head of a particular subclause should be its main content word, i.e. the nominal head. Parsers trained on UD Treebanks recognize dependency subtrees where the head is the semantic core of the subclause, effectively mitigating the dissociated nucleus phenomenon. We leverage this characteristic of UD parsers to automatically annotate the whole CoNLL-2009 corpus using `trankit` [16], which emerges as the strongest UD parser in the comparison we include in Appendix B.

### 3.1. SemDepAlign: subtree alignment

We introduce SemDepAlign, a novel algorithm for syntactic parse semi-alignment from the dependency annotations in CoNLL-2009 to UD, described in Algorithm 1. SemDepAlign is a deterministic subtree aligning algorithm that, for each role token $t$ associated with a predicate, finds the UD subtree that most closely matches the original subtree headed by $t$ in the original dependency tree of CoNLL-2009. It then returns the head node $t'$ of

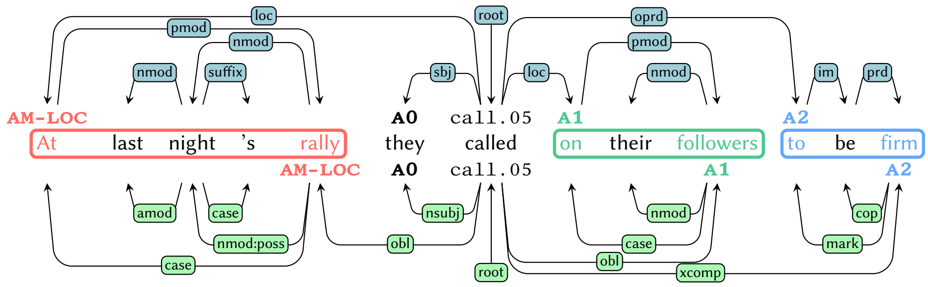


**Figure 1:** An edited example from the English CoNLL-2009 development set: original (top) and aligned (bottom) dependency and role annotations for the predicate *"called"*. We represent role annotations through colored clusters, where SemDepAlign aligns the head token to a more semantic token heading the UD subtree closest to the original.

the UD subtree, which will be assigned the role label in the aligned SRL annotation.

As shown in Algorithm 1, SemDepAlign starts from the UD root node (*root_ud*), loops over the nodes of the tree through a breadth-first search (BFS), and finds the node which heads the subtree with minimal symmetric set difference (*SymDiff*) between its tokens and the set of tokens in the original role span (*role_tokens*). The symmetric difference between two sets of tokens $S_1$ and $S_2$ is defined through the set operations *difference* (\) and *union* (∪) like so: $(S_1 \setminus S_2) \cup (S_2 \setminus S_1)$. Intuitively, if the symmetric difference between the original and the UD subtree is the empty set, they match exactly and we can simply select the head of the UD subtree as the role token. Otherwise, selecting the head of the UD subtree with the minimal symmetric difference compared to the original subtree is equivalent to selecting the subtree with the most overlap with the original span.

Figure 1 gives an example of the output of SemDepAlign: at the top of the figure we display the original annotation of the sentence derived from the English split of CoNLL-2009, with the presence of a dissociated nucleus in three of the four roles for the predicate "called"; in the bottom part we show the output of our alignment procedure, which moves the role annotations to the tokens that perform the semantic function.

### 3.2. Aligned-CoNLL09: analysis

We apply SemDepAlign to CoNLL-2009 to mitigate the dissociated nucleus phenomenon, obtaining the Aligned-CoNLL2009 dataset. After the application of SemDepAlign, the number of role token annotations that are modified is considerable over all CoNLL-2009 languages (between 21% and 32% of the total roles), except for Czech (~7%).

To gain a better understanding of the differences that the alignment process introduces, we consider the annotations of the original tokens that are modified by SemDepAlign and the resulting aligned role tokens. We measure three metrics on these two sets to evaluate their semantic richness:

- Number of **content words**, i.e. words that are either nouns, adjectives, adverbs, or verbs, which indicates that the heads identified by SemDepAlign are more varied (2713 vs. 680 for English, 3.99×);
- Number of **unique tokens**, which indicates that the heads identified by SemDepAlign are less repetitive (1906 vs. 477, 4×);
- Number of **unique synsets**, which indicates that the heads identified by SemDepAlign are associated with different meanings (1387 vs. 481, 2.88×) according to a Word Sense Disambiguation system [17].

From Table 1 we can see how SemDepAlign dramatically increases the semantic content of role tokens in English, Spanish and German, identifying more than 4× the number of content words, more than 2.5× the number of unique tokens and around 3× the number of unique synsets compared to the original annotations. We find a smaller but consistent increase of semantic content in Catalan and Chinese, whilst in Czech all metrics are similar, indicating a reduced effect of SemDepAlign.

## 4. Integrating re-associated nuclei

Although we demonstrate that re-associated nuclei in dependency-based SRL provide additional semantic information, an important research question is whether integrating our proposal into current systems can lead to a change in performance. Therefore, we build on top of the strong SRL model proposed by Conia and Navigli [18] and design a new approach that jointly learns both types of role annotations, i.e. the original role tokens and

| Language | Catalan | | Czech | | German | | English | | Spanish | | Chinese | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | O | A | O | A | O | A | O | A | O | A | O | A |
| # modified roles | 3356 (29.1%) | | 3578 (7.3%) | | 314 (26.9%) | | 4205 (30.3%) | | 3756 (32.4%) | | 4021 (21.7%) | |
| Content words | 1159 | 1344 | 2486 | 2923 | 59 | 246 | 680 | 2713 | 574 | 3019 | 595 | 618 |
| Unique tokens | 529 | 1952 | 2574 | 2473 | 108 | 274 | 477 | 1906 | 563 | 2324 | 1223 | 1742 |
| Unique synsets | 825 | 921 | 1339 | 1397 | 54 | 191 | 481 | 1387 | 457 | 1708 | 219 | 266 |

**Table 1**
Semantic variety of role tokens that were modified when aligning the original CoNLL-2009 (O) to Aligned-CoNLL09 (A).

| Lang | | Validation | | | Test | | |
|---|---|---|---|---|---|---|---|
| | | P | R | F1 | P | R | F1 |
| ca | B | 87.97 | 87.76 | 87.86 | 88.12 | 88.04 | 88.08 |
| | A | 87.46 | 87.01 | 87.23 | 87.16 | 86.88 | 87.02 |
| cs | B | 86.49 | 86.38 | 86.44 | 86.18 | 86.14 | 86.16 |
| | A | 86.42 | 86.44 | 86.43 | 86.19 | 86.20 | 86.20 |
| de | B | 90.52 | 90.72 | 90.62 | 89.82 | 90.26 | 90.04 |
| | A | 90.74 | 90.32 | 90.53 | 89.63 | 90.02 | 89.82 |
| en | B | 91.18 | 91.55 | 91.37 | 91.95 | 92.38 | 92.16 |
| | A | 91.38 | 91.33 | 91.35 | 92.07 | 92.25 | 92.16 |
| es | B | 86.79 | 86.92 | 86.85 | 86.20 | 85.65 | 85.93 |
| | A | 86.70 | 86.41 | 86.55 | 85.89 | 85.18 | 85.53 |
| zh | B | 89.46 | 88.97 | 89.22 | 89.47 | 88.81 | 89.14 |
| | A | 89.24 | 89.06 | 89.15 | 89.28 | 88.66 | 88.97 |

**Table 2**
Results on the validation and test sets of all languages in CoNLL-2009. 'B' indicates the baseline models' results, whilst 'A' indicates the results achieved by our aligned version.

| Lang | | Predicate F1 | Role F1 | Aligned Role F1 |
|---|---|---|---|---|
| ca | B | 98.79 | 87.45 | — |
| | A | 98.67 | 86.64 | 82.99 |
| cs | B | 99.38 | 89.55 | — |
| | A | 99.39 | 89.59 | 87.11 |
| de | B | 94.88 | 89.42 | — |
| | A | 94.55 | 89.64 | 86.42 |
| en | B | 95.15 | 89.75 | — |
| | A | 95.22 | 89.85 | 87.85 |
| es | B | 99.00 | 86.33 | — |
| | A | 98.99 | 85.57 | 81.93 |
| zh | B | 96.17 | 86.06 | — |
| | A | 96.05 | 85.88 | 83.15 |

**Table 3**
Finer-grained evaluation on all CoNLL-2009 test sets on predicates, roles and aligned roles. 'B' indicates the baseline models' results, whilst 'A' indicates the results achieved by our aligned version.

the aligned ones. In brief, this architecture derives a contextualized word representation for each word in a sentence from a BERT-like Pretrained Language Model [19, PLM]. It then applies a custom "fully-connected" stacked-BiLSTM sequence encoder to derive a predicate-aware representation, which is in turn used to derive a predicate- and argument-specific embedding for each word in the sentence. Finally, an argument-specific fully-connected BiLSTM is applied to further encode each word with respect to a specific predicate, from which it derives the final score distribution over the role vocabulary through a simple linear classifier. The model is trained to minimize the sum of categorical cross-entropy losses on predicate identification, predicate disambiguation and argument identification and classification.

To adapt this model for our joint modeling task, we duplicate the linear classifier for the semantic roles and set two different targets for the two role classifiers: the original role token and label from CoNLL-2009 and the aligned role token and label obtained with SemDepAlign. Our final loss adds terms for UD-aligned argument identification and classification to the original loss.

**Experimental setup** We use XLM-RoBERTa-base [20] as the underlying PLM, and leave other hyperparameters unchanged. We conduct our experiments on all of the language splits of CoNLL-2009, namely, Catalan, Czech, German, English, Spanish, and Chinese.

**Results** Table 2 compares the results of our joint-modeling alignment system against our baseline on the CoNLL-2009 validation and test sets. Importantly, we observe that the additional task of modeling the semantic core of an argument does not significantly alter the performance (very similar F1 score on the test), despite the added difficulty brought by the identification of semantic cores. Table 3, instead, provides a breakdown of the F1 scores on predicate, role and aligned role predictions. The aligned system is in line with the baseline despite being tasked with a more complex objective. More interestingly, we observe that the F1 score on the semantic heads is comparable, indicating that the model is able to identify UD-aligned roles effectively.

## 5. Semantic roles in AMR graphs

We also develop an evaluation method based on the Abstract Meaning Representation formalism [11, AMR] for Semantic Parsing. The interconnection between SRL and AMR is well-known across the literature [21, 22]: both

| Test dataset | Standard | Aligned | $\Delta$ |
|---|---|---|---|
| LORELEI | 65.98 | 71.33 | 5.35 |
| Weblog and WSJ | 64.07 | 70.13 | 6.06 |
| Xinhua MT | 67.92 | 75.68 | 7.76 |
| BOLT DF MT | 60.92 | 68.17 | 7.25 |
| BOLT DF English | 56.22 | 62.12 | 5.90 |
| Proxy reports | 24.50 | 22.40 | -2.10 |
| Average | 56.60 | 61.64 | 5.04 |

**Table 4**
AMR-precision metric over standard and aligned role predictions derived from test datasets in AMR3.0. $\Delta$ indicates the difference in precision between the unified roles and the standard ones.

tasks aim to construct a semantic representation of a sentence, although SRL, covering only surface-level semantic frames, is more superficial than AMR, which aims to provide a more complete and in-depth structured representation that can interconnect different semantic frames. Given that AMR aims to abstract away from the specific syntax of a sentence to focus only on its semantic content, our intuition is that a dependency-based SRL system is more "semantic" if its predictions of predicate-role pairs are contained in the AMR annotation for the same sentence.

Therefore, we devise the **AMR-precision** metric: given a sentence $S$, its golden annotated AMR graph $G_{\text{AMR}}$ with token-node alignments available and a set of dependency-based SRL predictions, we filter the predicted semantic frames so that the predicate of each frame is present in the golden AMR graph. We then compute the ratio between the number of role tokens that are connected to their predicate in the AMR graph over the total number of roles predicted.

Given the SRL system introduced in Section 4, we apply it to the AMR3.0 (LDC2020T02[1]) test datasets, keeping both the standard and the aligned role predictions. We then compute the AMR-precision for both sets of predicted roles, and compare them in Table 4. It is clear that aligned roles are more likely to be present in the corresponding AMR graph of a sentence, with a consistent difference in AMR-precision in all test datasets except *Proxy reports*. This particular dataset has a "templatic, report-like structure" as mentioned in the AMR3.0 guidelines, so it is possible that the reduced performance is due to this particular characteristic.

This finding can pave the way for future work exploring the linkage between these two fundamental semantic tasks, as also suggested in the multi-layer annotation provided in MOSAICo [23].

[1] catalog.ldc.upenn.edu/LDC2020T02

## 6. Related work

Syntactic information has always been considered important for recognizing semantic frames in SRL. Marcheggiani and Titov [24] were among the first to model the dependency information provided in dependency-based SRL, followed most recently by Xia et al. [25], Fei et al. [26]. These works differ in respect of modeling choices and in the kind of extra syntactic data to be included (e.g. constituency trees, POS tags).

We also considered other syntactic frameworks, such as HPSG [27], to align the role annotations. HPSG robustly models the relationship between semantic cores of a sentence, but the lack of automatic tools with an acceptable performance and the difficulty in aligning dependency-based subtrees to HPSG spans compelled us to use UD.

## 7. Conclusion

In this paper, we conducted an in-depth investigation on the dissociated nucleus issue in dependency-based SRL. We introduced SemDepAlign, a novel method to align predicate-argument structures in SRL with syntactic parses from the Universal Dependencies project, which addresses the dissociated nucleus phenomenon. Our analyses and experiments in SRL modeling demonstrate that our approach to dissociated nuclei brings more semantic richness whilst remaining competitive on standard benchmarks.

## 8. Limitations

A limitation of our work is that it builds upon existing dependency parsers trained on Universal Dependencies. These parsers have reached high robustness across many languages, between 85 and 93 in Labeled Attachment Score (LAS) on the languages present in CoNLL-2009. But the error that these automatic methods necessarily encounter propagates directly to our alignment algorithm, with no way of recovering from the mistake. This limitation would be even more impactful in languages where the automatic dependency parser performed worse, presumably in low-resource settings, preventing a robust expansion of our work to these settings.

A more methodological limitation of our contributions concerns the availability of the CoNLL-2009 dataset. Although it is a well-established corpus in the SRL literature, it has a proprietary licensing scheme and one must acquire the resource from the Linguistic Data Consortium (LDC). We trust that, given the importance of the corpus, this will not limit the relevance of our work.

## Acknowledgements

Simone Conia gratefully acknowledges the support of the PNRR MUR project PE0000013-FAIR, which fully funds his fellowship.

Roberto Navigli also gratefully acknowledges the support of the CREATIVE project (CRoss-modal understanding and gEnerATIon of Visual and tExtual content), which is funded by the MUR Progetti di Rilevante Interesse Nazionale programme (PRIN 2020).

## References

[1] D. Gildea, D. Jurafsky, Automatic Labeling of Semantic Roles, Computational Linguistics 28 (2002) 245–288. URL: https://doi.org/10.1162/089120102760275983. doi:10.1162/089120102760275983.

[2] L. Màrquez, X. Carreras, K. C. Litkowski, S. Stevenson, Semantic Role Labeling: An Introduction to the Special Issue, Computational Linguistics 34 (2008) 145–159. URL: https://doi.org/10.1162/coli.2008.34.2.145. doi:10.1162/coli.2008.34.2.145.

[3] Y. Zeng, X. Jin, S. Guan, J. Guo, X. Cheng, Event coreference resolution with their paraphrases and argument-aware embeddings, in: D. Scott, N. Bel, C. Zong (Eds.), Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 3084–3094. URL: https://aclanthology.org/2020.coling-main.275. doi:10.18653/v1/2020.coling-main.275.

[4] A. Marasović, A. Frank, SRL4ORL: Improving opinion role labeling using multi-task learning with semantic role labeling, in: M. Walker, H. Ji, A. Stent (Eds.), Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 583–594. URL: https://aclanthology.org/N18-1054. doi:10.18653/v1/N18-1054.

[5] M. Zhang, P. Liang, G. Fu, Enhancing opinion role labeling with semantic-aware word representations from semantic role labeling, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 641–646. URL: https://aclanthology.org/N19-1066. doi:10.18653/v1/N19-1066.

[6] S.-M. Kim, E. Hovy, Extracting opinions, opinion holders, and topics expressed in online news media text, in: M. Gamon, A. Aue (Eds.), Proceedings of the Workshop on Sentiment and Subjectivity in Text, Association for Computational Linguistics, Sydney, Australia, 2006, pp. 1–8. URL: https://aclanthology.org/W06-0301.

[7] J. Lawrence, C. Reed, Argument mining: A survey, Computational Linguistics 45 (2019) 765–818. URL: https://aclanthology.org/J19-4006. doi:10.1162/coli_a_00364.

[8] T. Falke, I. Gurevych, Utilizing automatic predicate-argument analysis for concept map mining, in: C. Gardent, C. Retoré (Eds.), Proceedings of the 12th International Conference on Computational Semantics (IWCS) — Short papers, 2017. URL: https://aclanthology.org/W17-6909.

[9] J. Hajič, M. Ciaramita, R. Johansson, D. Kawahara, M. A. Martí, L. Màrquez, A. Meyers, J. Nivre, S. Padó, J. Štěpánek, P. Straňák, M. Surdeanu, N. Xue, Y. Zhang, The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages, in: Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task, Association for Computational Linguistics, 2009, pp. 1–18. URL: https://aclanthology.org/W09-1201.

[10] L. Tesnière, Elements of Structural Syntax, John Benjamins, 2015. URL: https://www.jbe-platform.com/content/books/9789027269997.

[11] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, N. Schneider, Abstract Meaning Representation for sembanking, in: A. Pareja-Lora, M. Liakata, S. Dipper (Eds.), Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, Association for Computational Linguistics, Sofia, Bulgaria, 2013, pp. 178–186. URL: https://aclanthology.org/W13-2322.

[12] M. Surdeanu, R. Johansson, A. Meyers, L. Màrquez, J. Nivre, The CoNLL 2008 shared task on joint parsing of syntactic and semantic dependencies, in: CoNLL 2008: Proceedings of the Twelfth Conference on Computational Natural Language Learning, Coling 2008 Organizing Committee, Manchester, England, 2008, pp. 159–177. URL: https://aclanthology.org/W08-2121.

[13] M.-C. de Marneffe, T. Dozat, N. Silveira, K. Haverinen, F. Ginter, J. Nivre, C. D. Manning, Universal Stanford dependencies: A cross-linguistic typology, in: N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.), Proceedings of the Ninth

International Conference on Language Resources and Evaluation (LREC'14), European Language Resources Association (ELRA), Reykjavik, Iceland, 2014, pp. 4585–4592. URL: http://www.lrec-conf.org/proceedings/lrec2014/pdf/1062_Paper.pdf.

[14] M.-C. de Marneffe, C. D. Manning, J. Nivre, D. Zeman, Universal Dependencies, Computational Linguistics 47 (2021) 255–308. URL: https://aclanthology.org/2021.cl-2.11. doi:10.1162/coli_a_00402.

[15] J. Grimshaw, Argument Structure, The MIT Press, Cambridge, MA, 1990.

[16] M. V. Nguyen, V. Lai, A. P. B. Veyseh, T. H. Nguyen, Trankit: A light-weight transformer-based toolkit for multilingual natural language processing, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, 2021.

[17] R. Orlando, S. Conia, F. Brignone, F. Cecconi, R. Navigli, AMuSE-WSD: An all-in-one multilingual system for easy Word Sense Disambiguation, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021, pp. 298–307. URL: https://aclanthology.org/2021.emnlp-demo.34. doi:10.18653/v1/2021.emnlp-demo.34.

[18] S. Conia, R. Navigli, Bridging the gap in multilingual semantic role labeling: a language-agnostic approach, in: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 1396–1410. URL: https://aclanthology.org/2020.coling-main.120. doi:10.18653/v1/2020.coling-main.120.

[19] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[20] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 8440–8451. URL: https://aclanthology.org/2020.acl-main.747. doi:10.18653/v1/2020.acl-main.747.

[21] L. Chen, P. Wang, R. Xu, T. Liu, Z. Sui, B. Chang, ATP: AMRize then parse! enhancing AMR parsing with PseudoAMRs, in: M. Carpuat, M.-C. de Marneffe, I. V. Meza Ruiz (Eds.), Findings of the Association for Computational Linguistics: NAACL 2022, Association for Computational Linguistics, Seattle, United States, 2022, pp. 2482–2496. URL: https://aclanthology.org/2022.findings-naacl.190. doi:10.18653/v1/2022.findings-naacl.190.

[22] R. Navigli, Natural Language Understanding: Instructions for (Present and Future) Use, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, International Joint Conferences on Artificial Intelligence Organization, 2018, pp. 5697–5702. URL: https://doi.org/10.24963/ijcai.2018/812. doi:10.24963/ijcai.2018/812.

[23] S. Conia, E. Barba, A. C. Martinez Lorenzo, P.-L. Huguet Cabot, R. Orlando, L. Procopio, R. Navigli, MOSAICo: a multilingual open-text semantically annotated interlinked corpus, in: K. Duh, H. Gomez, S. Bethard (Eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 7990–8004. URL: https://aclanthology.org/2024.naacl-long.442. doi:10.18653/v1/2024.naacl-long.442.

[24] D. Marcheggiani, I. Titov, Encoding sentences with graph convolutional networks for semantic role labeling, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Copenhagen, Denmark, 2017, pp. 1506–1515. URL: https://aclanthology.org/D17-1159. doi:10.18653/v1/D17-1159.

[25] Q. Xia, R. Wang, Z. Li, Y. Zhang, M. Zhang, Semantic role labeling with heterogeneous syntactic knowledge, in: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 2979–2990. URL: https://aclanthology.org/2020.coling-main.266. doi:10.18653/v1/2020.coling-main.266.

[26] H. Fei, S. Wu, Y. Ren, F. Li, D. Ji, Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, Online, 2021, pp. 549–559. URL: https://aclanthology.org/2021.findings-acl.49. doi:10.18653/v1/2021.

| POS | Tag description | Frequency | Percentage (%) |
|---|---|---|---|
| NN | Noun, singular or mass | 111,931 | 23.18% |
| IN | Preposition or subordinating conjunction | 76,821 | 17.28% |
| NNS | Noun, plural | 46,256 | 10.40% |
| NNP | Proper noun, singular | 28,238 | 6.35% |
| VBD | Verb, past tense | 25,414 | 5.72% |
| VB | Verb, base form | 23,244 | 5.23% |
| VBN | Verb, past participle | 19,370 | 4.36% |
| JJ | Adjective | 18,308 | 4.12% |
| RB | Adverb | 17,423 | 3.92% |
| TO | to | 17,263 | 3.88% |
| PRP | Personal pronoun | 14,950 | 3.36% |
| VBG | Verb, gerund or present participle | 14,901 | 3.35% |
| VBZ | Verb, 3rd person singular present | 13,360 | 3.01% |
| MD | Modal | 9,316 | 2.10% |
| VBP | Verb, non-3rd person singular present | 7,774 | 1.75% |
| **Total** | | 475,069 | 100.00% |

**Table 5**
Frequency of POS Tags in the English split of CoNLL-2009.

findings-acl.49.
[27] C. Pollard, I. Sag, Head-Driven Phrase Structure Grammar, Studies in Contemporary Linguistics, University of Chicago Press, 1994. URL: https://books.google.it/books?id=Ftvg8Vo3QHwC.
[28] M. Straka, UDPipe 2.0 prototype at CoNLL 2018 UD shared task, in: Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 197–207. URL: https://aclanthology.org/K18-2020. doi:10.18653/v1/K18-2020.
[29] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, Stanza: A Python natural language processing toolkit for many human languages, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2020. URL: https://nlp.stanford.edu/pubs/qi2020stanza.pdf.

# A. Dissociated nuclei in non-English samples of CoNLL-2009

## A.1. Catalan

Original sentence:
"Piqué recomana les fusions entre empreses per millorar la rendibilitat."
Translation:
"Piqué recommends mergers between companies to improve profitability."
Dissociated nucleus:
In the clause "per millorar" ("to improve"), 'per' ('to') is tagged as `argM-fin` for predicate 'recomana' ('recommends') instead of the head of the subclause 'millorar' ('improve').

## A.2. German

Original sentence:
"Setzt Hessen auf eine Effizienzsteigerung der Verwaltung durch neue Steuerungsinstrumente ."
Translation:
"Hesse is focusing on increasing the efficiency of administration through new control instruments."
Dissociated nucleus:
Considering the predicate 'setzt' ('focus'), the clause "auf eine Effizienzsteigerung" ("on increasing the efficiency") is annotated with role `A1` on the token 'auf' instead of the semantic core 'Effizienzsteigerung'.

## A.3. Spanish

Original sentence:
"Don Antonio se encontraba en su casa cuando sonó el timbre de la puerta."
Translation:
"Don Antonio was at his home when the doorbell rang."
Dissociated nucleus:
The role "en su casa" ("at his home") for predicate 'encontraba' ('was') is tagged as `arg2-loc` on the token 'en' ('in') instead of the semantic nucleus 'casa' ('home').

## A.4. Chinese

Original sentence:
巴拉克 在 民意 测验 中 一直 表现 不 佳 。
Transliteration:
"Barak in public opinion test in continuously performance no good."
Translation:
"Barak has consistently underperformed in the polls."
Dissociated nucleus:
In the clause 在 民意 测验 ("in the public opinion polls") for the nominal predicate 佳 ('good'), as the token 在 ('in') is tagged as the `LOC` role, instead of the more semantic 测验 ('polls').

# B. Universal Dependency parsers

We consider three among the best off-the-shelf dependency parsers, namely, `trankit` [16], `UDPipe` [28] and `Stanza` [29]. Table 6 compares the reported evaluation of each parser on standard treebanks for Catalan, Czech, German, English, Spanish and Chinese. We choose `trankit` as it achieves a higher UAS and LAS than the two alternatives in all languages except Spanish (slightly worse than `UDPipe`), with a considerable margin in Chinese.

| Treebank | System | UAS | LAS |
|---|---|---|---|
| Catalan AnCora | `trankit` | **95.15** | **93.83** |
| | `UDPipe` | 94.92 | 93.43 |
| | `Stanza` | 93.55 | 91.66 |
| Czech PDT | `trankit` | **95.24** | **93.65** |
| | `UDPipe` | 95.01 | 93.64 |
| | `Stanza` | 92.22 | 90.18 |
| German GSD | `trankit` | **89.01** | **85.20** |
| | `UDPipe` | 87.04 | 83.20 |
| | `Stanza` | 85.80 | 81.80 |
| English EWT | `trankit` | **91.29** | **89.4** |
| | `UDPipe` | 90.71 | 88.81 |
| | `Stanza` | 88.90 | 86.77 |
| Spanish AnCora | `trankit` | 93.29 | 91.10 |
| | `UDPipe` | **93.68** | **91.92** |
| | `Stanza` | 93.09 | 91.30 |
| Chinese Simplified GSD | `trankit` | **87.38** | **84.82** |
| | `UDPipe` | 72.74 | 70.28 |
| | `Stanza` | 73.41 | 70.65 |
| **Average** | `trankit` | **91.89** | **89.67** |
| | `UDPipe` | 89.02 | 86.88 |
| | `Stanza` | 87.83 | 85.40 |

**Table 6**
Performance of multiple off-the-shelf dependency relation parsers, measured by the standard Unlabeled and Labeled Attachment Scores (UAS and LAS). Boldface scores indicate the best performing system on a specific treebank.