

ESTUDOS LINGUÍSTICOS E FILOLÓGICOS OFERECIDOS A IVO CASTRO

ESTUDOS LINGUÍSTICOS E FILOLÓGICOS

OFERECIDOS A IVO CASTRO



ISBN 978-989-98666-3-8

CENTRO DE LINGUÍSTICA DA UNIVERSIDADE DE LISBOA

2019

I,

En nunca guardei rebanhos,
Mas, e' como se o guardasse.
Quil' alma
~~Quil' alma~~ e' curba um pastor,
Pertence ao vento e ao sol,
E anda pela mar de Estacada
A carter e a hucar. (A viver como elles vivem.)
Toda a paz da natureza sem senti-
nem ventar - e a mesa lavr;
Mas em fin' tinto, por um e por coisa nenhuma,
Como um furo d'ol' para a alma magra,
Quando acontece ao fundo da planicie,
E' sent' que a vida degra um se ver
Como uma borlida ~~que se~~ de cura em cura.
- ja dentro de
Mas a mente tinto e' abgo
Logo e' natural e juto.

f. 19 GF-1
51

ESTUDOS LINGUÍSTICOS E FILOLÓGICOS
OFERECIDOS A IVO CASTRO

Organizados por

Ernestina Carrilho
Ana Maria Martins
Sandra Pereira
João Paulo Silvestre

CENTRO DE LINGUÍSTICA DA UNIVERSIDADE DE LISBOA

2019

Título • *Estudos linguísticos e filológicos oferecidos a Ivo Castro*

Organizadores • Ernestina Carrilho, Ana Maria Martins, Sandra Pereira, João Paulo Silvestre

Edição • Centro de Linguística da Universidade de Lisboa

Capa • Ivo Castro, arquivo particular

Contra-capas • Pormenor do manuscrito de Fernando Pessoa *Eu nunca guardei rebanhos*

[Alberto Caeiro], BN Esp. E3/67-1r. Biblioteca Nacional de Portugal.

ISBN 978-989-98666-3-8 [livro digital]

Índice geral

Prefácio	7
Comissão científica	13
<i>Laudatio</i>	
<i>Rosário Álvarez Blanco</i>	15
<i>A Estrada de Cintra</i> (Castro 2017): saudação aos participantes do IVº CILH	
<i>Inês Duarte</i>	29
Ivo Castro. Uma bibliografia (1969-2019).....	31
1. A demanda da ortografia mirandesa: entre a norma, a convenção e o florescimento	
<i>António Bárbolo Alves</i>	43
2. Os relativizadores <i>que</i> e <i>quem</i> em contextos preposicionados: análise de uma mudança do português clássico ao moderno	
<i>Aroldo de Andrade</i>	61
3. Periodização da história linguística do Sul da Bahia	
<i>Wagner Argolo Nobre</i>	89
4. Próclise e ênclise na oratória barroca	
<i>Ana Paula Banza</i>	119
5. A letra de samba – um <i>corpus</i> para estudos do português do Brasil no século XX	
<i>Flávio de Aguiar Barbosa</i>	141
6. Reflexos da paisagem humana e social da Idade Média em textos jurídicos: fenómenos de indireção ou de atenuação discursiva em atos diretivos	
<i>Clara Barros</i>	165
7. A variação entre <i>ter</i> e <i>haver</i> em construções existenciais numa variedade insular do PE (Funchal)	
<i>Aline Bazenga</i>	181
8. Subjuntivo <i>vs</i> indicativo em orações completivas: percurso diacrónico no português brasileiro	
<i>Rosane de Andrade Berlinck</i>	217
9. Continuando o debate sobre a origem do infinitivo flexionado: uma abordagem «mista» e baseada em <i>corpora</i>	
<i>Giulia Bossaglia</i>	245
10. <i>O artificio das etimologías</i> : elementos constitutivos da toponímia galega	
<i>Ana Isabel Boullón-Agrelo</i>	277
11. Estudo antroponímico dunha comunidade galega do século XV	
<i>Paula Bouzas</i>	321
12. Análise dialetométrica do Atlas Linguístico-Etnográfico da Região Sul do Brasil: variação lexical	
<i>Fernando Brissos, João Saramago</i>	349
13. Reformas ortográficas: práticas sem teorias	
<i>Luiz Carlos Cagliari</i>	381

14. O uso do subjuntivo ao longo da história do português <i>Dinah Callou, Érica Almeida</i>	393
15. Análise de <i>macrocorpora</i> e <i>microcorpora</i> para estudos de linguística histórica.....	415
16. Historiando o Português Brasileiro. Ivo Castro e o Projeto de História do Português de São Paulo <i>Juan M. Carrasco González</i>	431
17. Estruturas de focalização em peças portuguesas e brasileiras <i>Sílvia Regina de Oliveira Cavalcante, Maria Eugênia Lammoglia Duarte, Mayara Nicolau de Paula</i>	445
18. Fernando Pessoa 2.0: novas ferramentas para velhos problemas <i>Simone Celani</i>	461
19. A arbitrariedade de terceira pessoa em português brasileiro <i>Fernanda Cerqueira</i>	477
20. A gramaticalização de <i>Nossa Senhora</i> nos falares mineiro e fluminense <i>Bruna Amarante de Mendonça Cohen</i>	509
21. Coordenação de constituintes nominais com apenas um determinante em português europeu <i>Madalena Colaço, Carolina Gramacho</i>	521
22. Os sonetos que houve entre Antero de Quental e Oliveira Martins <i>Ângela Correia</i>	553
23. O estrato linguístico duocentista num manuscrito seiscentista - a <i>Vida de Santa Senhorinha de Basto</i> <i>Marta Louro Cruz</i>	569
24. Considerações acerca das palavras portuguesas terminadas em <i>-oulo</i> / <i>-oilo</i> <i>Przemysław Dębowski</i>	583
25. A edição crítica do teatro de espetáculo espanhol do século XVIII. Um caso prático nas comédias de Manuel Fermín de Laviano <i>Alberto Escalante Varona</i>	607
26. Do latim galaico ao galego-português. Processos de mudança <i>Maria Alice Fernandes</i>	637
27. Ascendentes textuais do Livro de Marco Polo <i>Maria Helena Garvão</i>	667
28. A codificação de tópico do sujeito nas construções participiais absolutas licenciadas em textos de autores portugueses dos séculos XIV, XV, XVI e XVII <i>Alba Verôna Brito Gibrail</i>	683
29. Porque é que os relógios não quebram os ponteiros em português europeu? <i>Anabela Gonçalves, Matilde Miguel</i>	713
30. Estereótipos e iconización nas representações do português em textos galegos da Idade Moderna <i>Ernesto González Seoane</i>	739
31. Pontes entre os crioulos portugueses de África e a história do português: um caso de estudo de /b/ e /v/ <i>Tjerk Hagemeijer</i>	765
32. Uma história escrita à mão: edição de documentos históricos brasileiros <i>Alicia Duhá Lose</i>	779

33. O papel da relação entre letra e música na investigação de elementos prosódicos em períodos passados da língua: análise de duas <i>Cantigas de Santa Maria</i> <i>Gladis Massini-Cagliari</i>	805
34. As preposições em português à luz da Linguística Cognitivo-Funcional <i>Vanda Cardozo de Menezes, Monclar Guimarães Lopes</i>	831
35. Carolina Michaëlis e Henry Lang: um diálogo entre romanistas <i>Lênia Márcia Mongelli, Yara Frateschi Vieira</i>	845
36. Variación scriptolingüística e tradición manuscrita da lírica trobadoresca: As variables <nh/n> e <ss/s> <i>Henrique Monteagudo</i>	859
37. Alçamento de vogais médias pretônicas no português brasileiro sul-rio-grandense: comportamento variável e retrato oitocentista <i>Roberto Francisco Nasi, Valéria Neto de Oliveira Monaretto</i>	961
38. Ecos de Gil Vicente (ca.1465–ca.1536) en la revista de vanguardia <i>1616 (English & Spanish Poetry)</i> <i>María Victoria Navas Sánchez-Élez</i>	989
39. Da transcrição como exercício de escolha múltiplo <i>Ariadne Nunes, José Camões</i>	999
40. A linguística histórica e o léxico diferencial: variação dialetal e sociolinguística de alguns regionalismos do Português falado na ilha da Madeira <i>Naideia Nunes Nunes</i>	1023
41. Intercompreensão em Línguas Românicas como ferramenta para a aproximação à História das Línguas <i>Francisco Javier Calvo del Olmo, Karine Marielly Rocha da Cunha</i>	1061
42. Valores sintáticos e semânticos com <i>ser</i> e <i>estar</i> no <i>Leal Conselheiro</i> de D. Duarte <i>Paulo Osório</i>	1085
43. Lexicologia e crítica textual: O <i>Vocabulário em Idioma Bengalla e Portuguez</i> e o <i>Marsden Lexicon</i> <i>Stephen Parkinson</i>	1103
44. Portugal, França e Brasil: uma, duas ou três críticas genéticas? <i>Carlota Pimenta</i>	1121
45. Os múltiplos valores do item <i>homem</i> em Português Antigo <i>Clara Pinto</i>	1147
46. Preposições complexas em Português: lexicalização e gramaticalização <i>José Pinto de Lima</i>	1181
47. Contacto e variação em caboverdiano: uma questão de tempo <i>Fernanda Pratas</i>	1213
48. De quanta filologia precisa um linguista e de quanta linguística precisa um filólogo <i>Maria Ana Ramos</i>	1239
49. Nomes deverbais não sufixados e nomes deverbais corradicais sufixados: condições de existência <i>Graça Rio-Torto</i>	1303
50. Para uma interpretação hermenêutica de uma edição crítico-genética: emendas de elocução e emendas de poética <i>Enrique Rodrigues-Moura</i>	1339

51. O apagamento das vogais átonas finais [i] e [o], diante de consoante e de pausa, a partir de dados do ALEPG: Barlavento algarvio e São Miguel <i>Maria do Carmo Sá Teles de Araújo Rolo</i>	1355
52. O futuro das humanidades digitais é o passado <i>Jorge Viana Santos, Cristiane Namiuti</i>	1381
53. Nomes portugueses, nomes de portugueses e nomes em português: norma linguística e mudança sociolinguística <i>João Paulo Silvestre</i>	1405
54. O legado germânico na antroponímia neológica do português do Brasil <i>Juliana Soledade, Mailson Lopes, Leticia Rodrigues</i>	1417
55. O texto que se lê de <i>O Seminarista</i> , de Bernardo Guimarães <i>Luana Batista de Souza</i>	1447
56. A propósito de uma pseudo «cantiga de amigo» provençal. Problemas linguísticos, exegeticos e atributivos <i>Giuseppe Tavani</i>	1471
57. Um olhar sobre a <i>Vita Christi</i> : proposta de filiação dos fragmentos da primeira parte da obra <i>Silvio de Almeida Toledo Neto</i>	1483
58. Assimilação vocálica, coloração e coalescência em sequências V_1V_2 na diacronia e na sincronia do português: uma proposta descritiva baseada na fonologia dos elementos <i>João Veloso</i>	1515
59. O léxico patrimonial no quinhentismo português <i>Fernando Venâncio</i>	1541
60. Phonological metrics in author identification and characterization <i>Marina Vigário, Carla Pires, Fernando Martins, Sónia Frota</i>	1561
Palavras-chave.....	1583

18. Fernando Pessoa 2.0: novas ferramentas para velhos problemas

Simone Celani

Sapienza Università di Roma

O Espólio Pessoa representa um exemplo extremo no âmbito da filologia, devido em particular à sua estrutura, muito ampla e complexa, mas que conserva poucos elementos da ordenação original deixada pelo autor. Na materialidade dos documentos originais estão presentes, porém, numerosos indícios que permitem reconstruir diferentes graus de proximidade genética entre as peças, dados fundamentais para editar criticamente as obras de Fernando Pessoa. Contudo, são inúmeros os elementos a ter em consideração, e conseguir uma visão global utilizando apenas os métodos tradicionais exigiria uma análise demasiado demorada. Na Universidade de Roma La Sapienza teve início um novo projeto experimental que visa examinar um amplo número de variáveis de uma forma automática, utilizando um algoritmo que desfruta as potencialidades dos Sistemas Artificiais Adaptativos (AAS, Artificial Adaptive Systems) e fornece um mapa gráfico das ligações entre elas (formalmente um MST, Minimum Spanning Tree). Nos últimos anos deu-se início a um primeiro teste baseado num *dataset* que contém uma formalização das variáveis materiais (suportes, canetas, etc.) das peças que compõem *O caso Vargas*, identificando uma série de famílias ou *clusters* em que as peças podem ser divididas, cada uma das quais possivelmente associada a uma fase da escrita da obra. O mapa resultante da análise automática (um exemplo é visível na Figura 1) pode ser confrontado com dados ligados aos conteúdos das peças e fornecer indicações úteis sobre as sincronias na produção de diferentes partes de uma mesma obra ou de obras diferentes. Ampliando o exame a obras mais extensas, a séries de obras ou, melhor ainda, ao Espólio na sua totalidade, será possível ter uma hipótese de visão estratigráfica do Espólio, de modo a identificar ligações entre partes diferentes da obra de Pessoa e a fornecer importantes indicações cronológicas e genéticas. Em conclusão, o que se pretende demonstrar não é, obviamente, a possibilidade de teorizar a existência de um filólogo automático ou virtual, mas a de integrar novas ferramentas no trabalho de um filólogo de carne e osso que tente enfrentar a difícil tarefa de reconstruir as fases da escrita da obra de Pessoa.

Palavras-chave: Espólio Pessoa, *Digital Humanities*, *Artificial Adaptive Systems*, filologia material, atribuição.

1. Introdução: (*Digital*) *Humanities*

Dantes, o que havia era incompreensão e ceticismo. As designadas *Digital Humanities* constituíam um campo de estudos marginal, para poucos iniciados, que não suscitava um interesse generalizado e era visto fundamentalmente com desconfiança. A aparente complexidade dos seus procedimentos era considerada uma

desculpa para envolver num manto de esoterismo os seus métodos, invalidando, na prática, os resultados obtidos. Hoje, pelo contrário, a situação parece ter-se invertido drasticamente, e tem-se a impressão de que não existem projetos de investigação na área humanística que não incluam o recurso às novas tecnologias. A difusão dos instrumentos digitais aumentou exponencialmente, mesmo entre o público não especializado, graças ao acesso cada vez mais simples e económico a *software*, muitas vezes gratuito, bem como à abundância e difusão em larga escala de *hardware* capaz de o fazer funcionar.

Esta situação pode ser interpretada de dois modos: ou como um triunfo, o da chegada da idade do ouro das *Digital Humanities*, em que finalmente foi reconhecido o valor pragmático e metodológico desta nova abordagem, ou, mais melancolicamente, como a explosão de uma moda, um momento de euforia em que os meios amiúde se confundem com os fins da investigação, e proliferam projetos aparentemente inovadores e prometedores, mas que na realidade se limitam a rodar em torno de si mesmos, «a entreter a razão», como o «comboio de corda» da *Autopsicografia* de Pessoa. É mais ou menos esta a posição expressa por Lorenzo Tomasin, a 7 de julho de 2016, nas páginas do jornal *Sole 24 Ore*, num artigo sugestivamente intitulado «Umanisti scann(erizz)ati», que conclui o seguinte:

*A pratiche antiche come la lettura, l'analisi, la discussione, si preferiscono tecniche quali la scansione, la visualizzazione, l'automazione, che relegano alla fruizione meramente aneddotica contenuti del tutto intercambiabili, scelti solo a motivo della loro gradevolezza, popolarità o mediatica attrattività. Trasformati, insomma, da fini in mezzi occasionali. Pur evitando di cadere in indebite generalizzazioni o in eccessi di reazione, si tratta pur sempre di una tendenza della quale è bene stare in guardia*¹⁷¹.

Se é verdade que a primeira das duas interpretações acima expostas é certamente demasiado otimista, a segunda talvez não esteja muito longe da realidade, não fosse a circunstância de ignorar uma consequência menos visível, mas ainda assim importante dessa revolução (efetiva ou aparente): o facto de um número não negligenciável de pessoas não especialistas na área terem tido acesso a novos instrumentos, muitas vezes criados para fins bastante diferentes dos seus, que não se tornaram objeto único dos

¹⁷¹ O conceito (como, aliás, parte do texto, reformulado) encontra-se expresso em diversos passos da obra de Tomasin (2017). O artigo original pode ser consultado no seguinte endereço http://www.ilsole24ore.com/art/cultura/2016-07-07/umanisti-scannerizzati-113515.shtml?uid=ADa6Ivn&refresh_ce=1

seus estudos, mas que, pelo contrário, se prestam mais ou menos casualmente a facilitar um percurso já antes iniciado, a responder a perguntas para as quais até àquele momento era deveras trabalhoso dar resposta.

Nasce aqui uma nova dimensão interdisciplinar, em que problemas de natureza muito diversa encontram um denominador comum metodológico, e especialistas de matérias muito distantes podem encontrar instrumentos em comum, embora reinventados em cada caso para necessidades completamente novas. Se um filólogo, para efeitos de atribuição, vê utilidade num algoritmo originalmente criado para a compressão de dados, ou um arqueólogo inventa um uso não convencional para as redes neurais, não se trata de meros exercícios inúteis, de aparência robotizante, destinados a atrair financiamentos, mas de tentativas concretas de responder a problemas antigos através de novos pontos de vista, aplicando inversões de perspetiva e unindo dados e questões provenientes de disciplinas que podem ser muito diferentes. No fundo, o que a melhor investigação sempre fez.

É certamente verdade que, por trás desta extrema difusão de novos métodos, por trás desta espécie de corrida ao ouro digital, haverá muito para eliminar. Mas é igualmente verdade que se trata de um período pioneiro, não tanto de métodos, como de instrumentos. Porque é apenas disto que se trata, de instrumentos, de meios, e não de respostas. O que interessa continua a ser o mesmo objeto de antes: o texto, a obra, o dado arqueológico, o homem.

Naturalmente, é ainda demasiado cedo para fazer qualquer balanço do fenómeno das *Digital Humanities*. Deseja-se apenas que, num determinado momento, elas se integrem na área dos estudos humanísticos de modo tão completo e produtivo, que se perca a necessidade da presença do termo *Digital* inicial, que não é de facto o que importa realmente, ou o que importará. Visto que os próprios instrumentos digitais, no fim de contas, foram, também eles, criados pelo homem.

2. O Espólio Pessoa

Passados mais de oitenta anos sobre a morte de Fernando Pessoa, o seu principal legado, representado pelos seus documentos originais, continua a ser objeto de constantes pesquisas, que estão longe de terem atingido os seus objetivos. Está ainda distante a inteligibilidade completa dos dados neles contidos, devido a uma história

complexa e que nem sempre respeita o seu valor, que seria escusado repetir aqui pela enésima vez¹⁷². Muitos vestígios originariamente presentes, partindo logo da ordenação deixada pelo autor, ficaram irremediavelmente perdidos, e a filologia várias vezes envidou esforços para reconstruir obras cujos vestígios estão dispersos num espólio de 27 545 peças (Castro 2013: 89–95), para não contar o que lamentavelmente foi subtraído ao acervo principal, conservado há décadas na Biblioteca Nacional de Lisboa¹⁷³. Além dos fundamentais indícios presentes no conteúdo, a ecdótica tem escrutinado também os dados materiais, avançando importantes hipóteses sobre a cronologia dos documentos e as estratificações internas do Fundo. Não obstante isto, em muitos casos, os resultados obtidos mantiveram-se num nível de probabilidade que não consente respostas definitivas. Neste sentido, a implementação das novas possibilidades oferecidas pelos instrumentos digitais, não podendo claramente ser considerada resolutiva, representa pelo menos um campo a explorar com atenção.

Foi já por outros apontado quanto uma representação intertextual e digital poderia contribuir para a fruição das edições críticas das obras de Pessoa¹⁷⁴, e estão já em curso algumas experiências nesse sentido, como por exemplo o *Arquivo LdoD* de Portela e Rito Silva (2017), assim como exemplos de integração entre suporte em papel e suporte digital (Pessoa 2015). Nesta ocasião pretende-se falar de outro aspeto, que é o da utilidade de algumas novas metodologias de investigação na interrogação dos documentos, no tratamento do vasto número de dados extraíveis do Fundo, na análise detalhada dos originais na sua materialidade, na identificação de relações genéticas até agora não consideradas.

Nos últimos anos, na Universidade de Roma La Sapienza, tomaram forma de modo independente várias linhas experimentais de investigação, podendo cada uma das quais encontrar uma resposta para um aspeto diferente do enigma pessoano. Trata-se dos projetos *Archeosema*, iniciado em 2011 por Marco Ramazzotti, arqueólogo especializado no Próximo Oriente Antigo; *THESMA* (Terahertz and Spectrometry Manuscript Analysis), dirigido por Paola Italia, filóloga italiana, em colaboração com alguns investigadores da Faculdade de Física; e da investigação sobre o atribucionismo automático, coordenada por Paolo Canettieri, filólogo românico.

¹⁷² É possível encontrar uma síntese da questão em Celani (2005: 9–18).

¹⁷³ Espólio E3 do Arquivo de Cultura Portuguesa Contemporânea da BNP.

¹⁷⁴ Por exemplo Castro (2013: 117).

3. Archeosema

Archeosema é um projeto interdisciplinar que visa a aplicação à área humanística das metodologias ligadas aos Sistemas Adaptativos Artificiais (SAA), financiado pela Universidade La Sapienza em 2011 e dirigido por Marco Ramazzotti¹⁷⁵. O projeto, cuja denominação completa é «Archeosema. Sistemi informativi geografici e Sistemi Artificiali Adattivi per la ricerca teorica, applicativa e sperimentale di fenomeni complessi»¹⁷⁶, representa uma abordagem mais metodológica do que disciplinar, e visa verificar a utilização dos SAA para fins preditivos, de *datamining* ou para oferecer hipóteses de ordenação de dados em campos muito diversos, tais como a arqueologia, a geografia, a estatística, a linguística e a filologia.

O algoritmo da rede neural artificial (designado *AutoContractive Map* ou, abreviando, *AutoCM*¹⁷⁷) foi fornecido pelo Centro de Investigação «Semeion» de Paolo Massimo Buscema¹⁷⁸ e aplicado aos casos particulares em estudo graças ao apoio de um investigador do centro, Massimiliano Capriotti. Os resultados da investigação confluíram na publicação de um suplemento da revista *Archeologia e Calcolatori*, que continha uma ampla apresentação do projeto, das suas premissas teóricas e das aplicações e experiências efetuadas nas várias disciplinas (Ramazzotti 2013).

Em todos os casos de estudo analisados, havia dados parciais sobre objetos para os quais se pretendia reconstituir a forma original perdida: no caso da arqueologia, topografias parcialmente reconstituíveis através de escavações arqueológicas, ou fragmentos de placas de terracota contendo informações de arquivo; no caso da filologia, documentos incompletos e, muitas vezes, sem indicações claras de atribuição a obras específicas, dispersos num vastíssimo arquivo de autor do qual se perdeu a ordenação original. A aplicação do *AutoCM* nestes casos tinha o objetivo de obter hipóteses de ordenação e/ou preenchimento de lacunas existentes, de forma coerente e a partir dos dados conhecidos. Estas hipóteses teriam, posteriormente, de ser

¹⁷⁵ De entre a ampla bibliografia sobre a relação entre AAS e métodos arqueológicos, cfr. pelo menos Ramazzotti (2010 e 2013).

¹⁷⁶ «Archeosema. Sistemas informativos geográficos e Sistemas Adaptativos Artificiais para a investigação teórica, aplicada e experimental de fenómenos complexos».

¹⁷⁷ Sobre a qual, cfr. Buscema (2008).

¹⁷⁸ <http://www.semeion.it/>.

verificadas na prática ou por aplicação do processo a casos de estudo já resolvidos mediante outras metodologias.

O procedimento aplicado em todos os casos previa um mesmo percurso, que pode ser sumariamente descrito em quatro etapas: 1) tradução do *dataset* em matriz binária; 2) introdução do produto da tradução no software com o *AutoCM*; 3) criação de uma *minimum spanning tree* (MST) com base no *output* do *AutoCM*; 4) visualização gráfica da MST através do software *open source* Gephi¹⁷⁹. O grafo resultante da MST foi o principal instrumento hermenêutico utilizado (Figura 1), uma vez que permitiu a visualização sintética (mas não simplificadora) do problema complexo e não linear que se pretendia estudar.

O procedimento foi testado, no meu caso, num *dataset* não muito extenso, mas complexo devido ao número de variáveis incluídas (tratava-se de uma matriz de 184 x 95, 184 registos com 95 variáveis cada um) contendo dados relativos à descrição material (tipos e características dos suportes, instrumentos de escrita, etc.¹⁸⁰) de um pequeno conjunto de documentos do Fundo, sobre os quais tinha trabalhado alguns anos antes para uma edição que preparei¹⁸¹.

A obra em causa, intitulada *O caso Vargas*, é uma *detective story* vasta mas incompleta, escrita por Pessoa ao longo de um período de pouco mais de uma década, pertencente a uma série policial por ele idealizada e intitulada *Quaresma, decifrador*,

¹⁷⁹ <http://gephi.github.io/>.

¹⁸⁰ Para cada folha do Fundo Pessoa contendo fragmentos de O Caso Vargas, foram indicadas variáveis de «conteúdo» (eventual título do fragmento, eventual atribuição a um capítulo da obra, género literário, eventual atribuição ao autor fictício, língua), «suporte» (largura e comprimento da folha, presença de eventuais marcas de água, cor, eventuais cabeçalhos, eventuais indicações de número de página, se a folha estava inteira ou tinha sido cortada, dobrada ou não, furada ou não), «instrumentos de escrita» (se escrito à máquina ou manuscrito, no segundo caso se a caneta ou a lápis, cor da tinta), «data» (eventual data indicada no documento, datação resultante de outros dados materiais, pontual ou, caso contrário, se possível, *post e/ou ante quem*, eventual data de publicação dos conteúdos do documento), «documentos relacionados» (outros documentos do Fundo com os quais se relaciona diretamente o documento em questão, através de dados materiais ou de conteúdo, e eventuais outros documentos do Fundo contendo cópias ou outras redações do mesmo texto). Cada variável, na tradução em matriz binária do *dataset*, foi dividida em várias variáveis, de modo a poder ser definida sempre com as alternativas 0/1, sim/não, de uma condição específica. Por este motivo, muitas variáveis se multiplicaram (pode-se dar um exemplo com a cor das tintas: neste caso, de uma variável única com três alternativas possíveis – preto, azul e vermelho –, chegou-se a três variáveis – preto, azul e vermelho –, devendo indicar-se para cada uma delas uma condição positiva ou negativa, com resultados diferentes, tendencialmente dois 0 e um 1, mas por vezes também dois ou três 1, no caso de documentos escritos com canetas e tintas diferentes). Tudo isto explica que se tenha chegado a um número total tão alto de variáveis: 95.

¹⁸¹ Pessoa (2006), edição divulgativa e traduzida para o público italiano, cujas premissas filológicas foram apresentadas na minha tese de doutoramento, ainda parcialmente inédita (Celani 2004).

a partir do nome do investigador protagonista (Abílio Quaresma). A análise dos documentos originais que contêm o texto, efetuada recorrendo aos instrumentos tradicionais da filologia de autor, indica que a obra foi escrita entre o início dos anos Vinte e o ano da morte de Pessoa (1935), num processo que não foi contínuo nem sistemático. Existem esquemas preparatórios que apontam uma estrutura que, no seu último estágio de elaboração, devia ser composta por quinze capítulos, dois dos quais nunca chegaram a ser escritos, enquanto os outros se encontram em diferentes etapas de concretização. As características materiais dos documentos sugerem uma escrita dividida em diferentes fases, com uma elaboração que se foi ampliando como que em manchas irregulares: não linearmente, do princípio ao fim, mas acrescentando progressivamente excertos pertencentes a diferentes partes do texto. Alguns capítulos muito vastos foram dedicados à análise do processo de investigação praticado por Quaresma e à descrição psicológica do assassino, e apresentam, em grande parte, características quase ensaísticas, o que permite imaginar que podem ter sido escritos, em parte, independentemente da narrativa, e só depois integrados nela.

O grafo resultante desta experiência (Figura 1) fornece, como procurarei demonstrar, indicações úteis para identificar uma possível cronologia das fases de escrita da obra em causa¹⁸².

¹⁸² A descrição analítica da experiência e dos seus resultados encontra-se em Celani (2014).

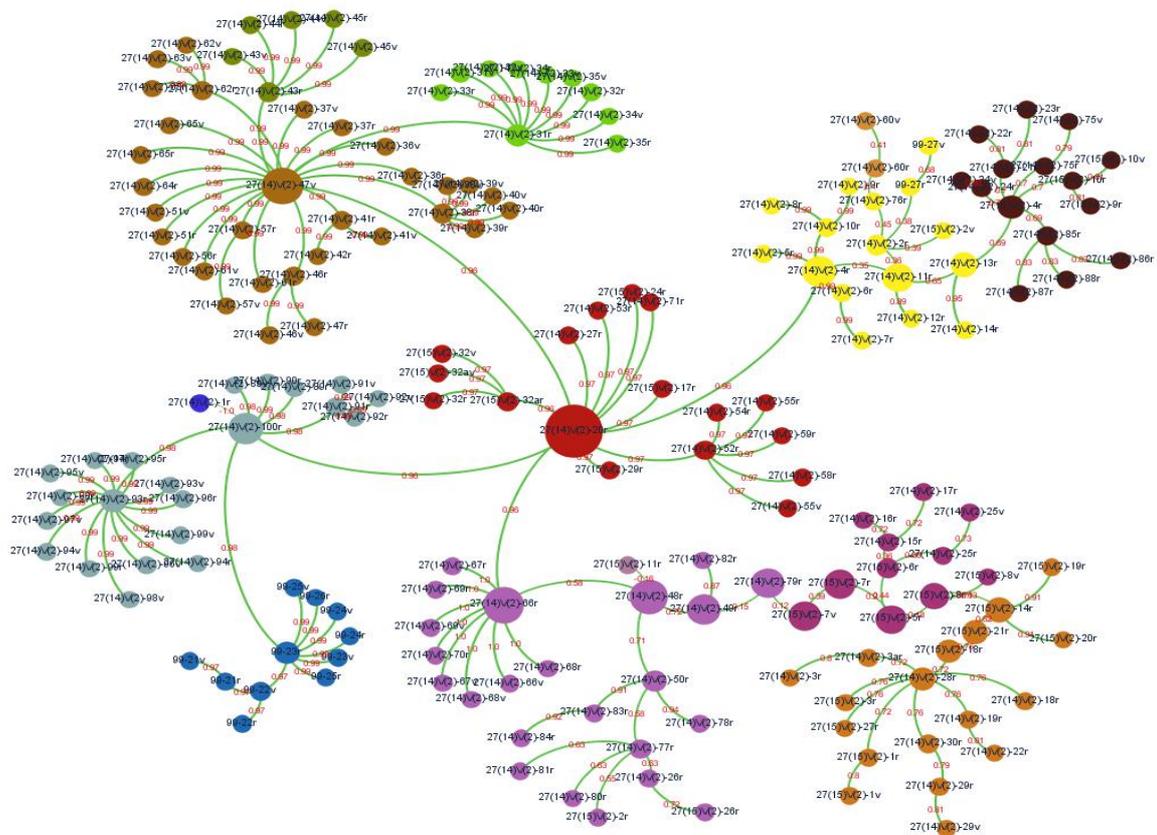


Figura 1. Grafo do MST que organiza os fragmentos de Caso Vargas de Fernando Pessoa a partir da elaboração fornecida pelo AutoCM

O *AutoCM* produziu, como *output*, uma série de valores que indicam a força da ligação entre cada documento que compõe o *Caso Vargas* com cada um dos outros documentos do *dataset*, tomando em consideração todas as variáveis envolvidas e pondo em evidência as coincidências entre as características materiais de cada um deles. A hipótese avançada é de que uma coincidência material (não só de uma, mas da maior parte das variáveis envolvidas) corresponda também a uma proximidade genética. O grafo resultante do processamento do *AutoCM* mostra, portanto, os vários documentos que compõem o *Caso Vargas*, fornecendo a representação mais económica das ligações materiais presentes entre eles e indicando o respetivo peso (numa escala que vai de um mínimo de -1 a um máximo de $+1$). Na prática, isto

permite visualizar os documentos em grupos materialmente homogêneos e, logo, provavelmente incluídos na mesma fase cronológica de gênese do texto, obtendo, em conclusão, uma representação gráfica das prováveis campanhas de escrita da obra.

Na complexidade material do Fundo Pessoa, este elemento não fornece apenas indicações genéticas, mas também editoriais, justamente porque, como se disse, muitas vezes os documentos deixados pelo grande escritor português são fragmentários e incompletos, e nem sempre é possível reconstituir facilmente a estrutura de uma obra, porque os fragmentos individuais não contêm indicações explícitas suficientes nesse sentido.

O processo de escrita do *Caso Vargas*, acima descrito, que envolveu um trabalho simultâneo em diferentes partes do texto e o aproveitamento de muito material reutilizado (o que parece caracterizar a escrita de muitas outras obras de Pessoa, se não todas, ou quase) parece confirmado pela MST elaborada pelo *AutoCM* (Figura 1), onde é possível identificar cinco *clusters* (um central e quatro periféricos) que, se confrontados com os elementos de datação direta de que dispomos, podem ser associados às diferentes fases de escrita do texto. Em particular, o *cluster 5* (o que se vê na parte inferior direita da figura) é o menos homogêneo, com ligações de peso mediantemente muito mais limitadas do que os outros, e talvez não seja por acaso que coincide com os tais fragmentos «ensaísticos» acima referidos.

Diga-se, no entanto, que os resultados obtidos só se prestam parcialmente a uma verificação através das metodologias tradicionais, e que a sua validade carece de garantias devido à ausência, no caso de Pessoa, de problemáticas semelhantes já totalmente resolvidas que permitam «testar» o sistema. A única possibilidade de que dispomos, neste caso, é a comparação dos resultados obtidos pelo processamento do *AutoCM* com as conclusões oferecidas independentemente por outras metodologias. É evidente, porém, que estamos assim limitados à esfera da probabilidade, e por isso a experiência descrita possui atualmente um valor experimental, de exploração. Trata-se de uma experiência com uma validade possível, e não certa, que deverá ser corroborada por muitas outras experiências antes de ser considerada efetivamente válida. De resto, contrariamente ao que acontece nas ciências exatas, nos estudos humanísticos (ou, se preferirmos, nas «ciências humanísticas»), a obtenção de um resultado definitivo, ainda que limitado ou parcial, é sempre necessariamente uma

utopia longínqua. É aqui que reside um dos elementos fundamentais de distância entre as «duas culturas». O que não invalida que alguns dados, mesmo nos estudos humanísticos, possam ser quantificáveis e, portanto, passíveis de análise recorrendo a metodologias típicas das designadas «ciências duras».

4. THESMA Project

O projeto *THESMA*¹⁸³ é outro estudo financiado pela Universidade La Sapienza, chefiado por Paola Italia, filóloga italiana atualmente em serviço na Universidade de Bolonha, que beneficiou da colaboração entre filólogos (de âmbito clássico, moderno e contemporâneo) e físicos para o desenvolvimento de uma máquina que permita efetuar uma análise espectrométrica de manuscritos literários, a fim de recolher dados materiais não acessíveis por outros métodos de observação. A análise permite, em particular, ter acesso a partes de texto apagadas e ler por baixo de tiras de papel ou entre páginas coladas:

A partire dagli anni Novanta, sono state varie le metodologie – dalla riflettografia infrarossa alla spettrometria attraverso i raggi NIR/VIS/ UV e X, alla fluorescenza – utilizzate per lo studio di manufatti artistici (pittorici, scultorei, o reperti archeologici) e, in alcuni casi, anche di pergamene, come per esempio all'interno del progetto Re.Co.Rd (Recupero Conservazione Ripristino Digitale) attivato dalla Biblioteca Medicea Laurenziana di Firenze. Queste nuove tecnologie non sono tuttavia ancora state applicate intensivamente a manoscritti antichi e moderni (VI-XXI secolo), nonostante la rilevanza dei potenziali casi di studio – dal Codice terenziano postillato da Pietro Bembo, al Codice degli abbozzi di Petrarca, dai manoscritti dei Canti di Leopardi o dei Promessi Sposi manzoniani, alle carte alluvionate di Carlo Emilio Gadda conservate presso l'Archivio Alessandro Bonsanti del Gabinetto Vieusseux di Firenze – sui quali potrebbero essere fruttuosamente sperimentate. Gli studi di filologia classica e italiana, in cui la scuola italiana è particolarmente avanzata, procedono ancora utilizzando riproduzioni tradizionali, come microfilm, in alcuni casi digitalizzati, oppure riproduzioni digitali ad alta definizione, che tuttavia, pur migliorando la qualità della resa grafica, non usufruiscono appieno di queste tecniche. In particolare, la filologia d'autore praticata in Italia ha trovato opportuna applicazione nella identificazione e nello studio degli strati correttori, che, con la pratica diffusa delle riproduzioni digitali, hanno cambiato radicalmente le procedure di analisi, rappresentazione e studio critico delle varianti. Le riproduzioni ad alta definizione, tuttavia, per quanto analitiche (superiori ai 5 mega), non permettono di indagare gli strati correttori, né, in casi particolarmente complessi, di individuare le varianti sotto cassatura e non hanno permesso di risolvere molti celebri casi di studio. L'applicazione di tecniche di rilevazione avanzata, unite all'analisi filologica dei manoscritti, ha portato alla decifrazione, seriazione diacronica, formalizzazione, interpretazione dei manoscritti d'autore a un livello di specializzazione di portata

¹⁸³ Está disponível uma síntese do projeto no endereço <http://www.filologiadautore.it/wp/thesma-project-sapienza-ricerca-2014-2016/>; cf. também Bonsi, Del Re, Italia, Ortolani (2016).

internazionale. In particolare, l'uso di tecniche di rilevazione avanzata risulta di forte impatto scientifico nei seguenti ambiti: 1. RILEVAMENTO degli strati correttori mediante l'individuazione dello spettro visivo dei diversi pigmenti dell'inchiostro con cui sono state apportate le serie correttorie; 2. RAPPRESENTAZIONE digitale dei singoli strati, che possono essere individualizzati, restituendo così online la lezione base, separata dai singoli interventi nelle loro successive stratificazioni; 3. STUDIO delle stratigrafie correttorie, per la costituzione del testo critico, per la rappresentazione in fasce di apparato e la resa digitale delle varianti; 4. CONSERVAZIONE dei materiali in un supporto digitale che possa sostituire e anzi preferirsi all'originale per una migliore resa delle caratteristiche tecniche del supporto cartaceo e del testo (Bonsi 2016: 153–154).

No que respeita ao âmbito específico pessoano, a experiência está, por enquanto, apenas no papel, uma vez que o protótipo «portátil» da máquina ainda não foi construído (Bonsi 2016: 154–155), tendo sido o projeto temporariamente suspenso por falta de fundos (mas uma linha de pesquisa, tecnicamente mais avançada, está a ser desenvolvida¹⁸⁴). Dado que os originais do Fundo Pessoa não podem, evidentemente, sair da secção de Reservados da Biblioteca Nacional de Portugal, de momento a experiência está adiada, mas há ensaios efetuados com manuscritos originais de Emilio Gadda que forneceram resultados encorajantes (Italia 2017: 47–56).

No meu entender, é também evidente que uma articulação entre os dados materiais obtidos no projeto THESMA e os instrumentos implementados no projeto Archeosema poderia trazer consequências muito relevantes para o âmbito da filologia pessoana.

5. Atribuicionismo automático

A partir de 2005, uma equipa chefiada pelo filólogo românico Paolo Canettieri, da Universidade de Roma La Sapienza, tem trabalhado no desenvolvimento de instrumentos que permitam automatizar a resolução de problemas de atribuição em âmbito literário e não só. Sobre as premissas teóricas do projeto, vejam-se alguns artigos já publicados (Canettieri 2012, 2016 e 2018). Limitar-me-ei agora a falar brevemente de como foi conduzida a experiência no caso de Pessoa e da possível relevância que estes métodos podem apresentar no âmbito dos problemas ecdóticos e críticos ligados à sua obra¹⁸⁵.

¹⁸⁴ <http://www.filologiadautore.it/wp/erc-ad-2017-3dthesma-project/>

¹⁸⁵ A parte que se segue consiste na reformulação parcial da informação já apresentada em Canettieri (2018).

As experiências foram todas realizadas seguindo uma única sequência de procedimentos comum. Em cada um dos casos, foi confrontado o mesmo número de versos de Pessoa (ortónimo e heterónimo) e de um outro autor, partindo de poetas cronológica e estilisticamente distantes e aproximando-se gradualmente à época e às características textuais das obras de Pessoa. Os textos foram divididos em cinco blocos textuais para cada autor. O número standard de versos para cada bloco textual foi estabelecido em 56, correspondente a 4 sonetos. Nos casos em que os autores produziram sonetos, optou-se por inserir textos pertencentes a este género, para manter tanto quanto possível a homogeneidade a nível quantitativo/textual (número de versos, mas também número de sílabas ou caracteres), no caso contrário, foram selecionados blocos de extensão equivalente, na medida do possível. Os autores selecionados são todos eles figuras que, com certeza, Pessoa conhecia extremamente bem e que fazem parte sempre das suas influências, a partir de Luís de Camões, passando por Antero de Quental e Camilo Pessanha, para chegar por fim a Mário de Sá Carneiro.

Em cada experiência, deram-se instruções ao *software* para atribuir um bloco textual a um dos dois autores; de cada vez, havia quatro ficheiros para cada um dos dois autores a funcionar como textos de comparação conhecidos, e os dois restantes ficheiros (um por autor) tinham de ser atribuídos a um ou ao outro. Todos os ficheiros, reconhecidos e por reconhecer, foram sendo inseridos até se obterem todas as combinações possíveis.

Num primeiro conjunto de experiências (da primeira à quarta) confrontaram-se estes quatro autores com uma seleção de sonetos de Pessoa ortónimo. Num segundo conjunto (da quinta à oitava), confrontaram-se os mesmos textos dos mesmos autores com uma seleção de poesias selecionadas da produção do ortónimo e dos heterónimos (Alberto Caeiro, Ricardo Reis, Álvaro de Campos). A última experiência, a única que envolvia um teste de atribuição dentro da obra de Pessoa, confrontou poesias de Pessoa ortónimo e do heterónimo Álvaro de Campos, mantendo o número de blocos e o número de versos por bloco dos testes anteriores. A tabela seguinte sintetiza os primeiros resultados obtidos:

Teste n.º	Tipo de atribuição	Autor 1	Autor 2	% de reconhecimento
1	Externa	Pessoa ortónimo (sonetos)	Luís de Camões (sonetos)	100
2	Externa	Pessoa ortónimo (sonetos)	Antero de Quental (sonetos)	100
3	Externa	Pessoa ortónimo (sonetos)	Camilo Pessanha (sonetos)	100
4	Externa	Pessoa ortónimo (sonetos)	Mário de Sá Carneiro (poesias)	100
5	Externa	Heterónimos de Pessoa (poesias)	Luís de Camões (sonetos)	100
6	Externa	Heterónimos de Pessoa (poesias)	Antero de Quental (sonetos)	100
7	Externa	Heterónimos de Pessoa (poesias)	Camilo Pessanha (sonetos)	100
8	Externa	Heterónimos de Pessoa (poesias)	Mário de Sá Carneiro (poesias)	100
9	Interna	Pessoa ortónimo (poesias)	Álvaro de Campos (poesias)	70

O resultado das experiências acima descritas permite salientar pelo menos duas ideias, provavelmente banais, mas nem sempre consideradas de modo correto quando se fala de Pessoa, pelo menos a nível crítico (ainda que muito claras para os filólogos pessoanos): 1) que Pessoa é *um* autor, identificável e reconhecível, apesar do seu esforço constante de diferenciação estilística e de conteúdos, concretizado no sistema dos heterónimos; 2) que o sistema de atribuição automático utilizado não pode ser útil para resolver problemáticas de atribuição interna, porque o «disfarce» adotado por Pessoa por meio dos heterónimos não chega ao ponto de «desnaturar» completamente as características que o identificam como autor, mas funciona perfeitamente para resolver problemas de atribuição externa, semelhantes aos que já surgiram ao longo da história editorial das suas obras.

6. «P'ra quê?»: uma primeira tentativa de conclusão

Para concluir, gostaria de terminar, de forma aparentemente paradoxal, com uma questão que, na realidade, era necessário colocar no início da presente exposição, e não no fim. Justificar-me-á, por ventura, o facto de não pretender, nesta ocasião, chegar a nenhuma conclusão definitiva, mas tão somente sugerir a ideia de que nas experiências aqui descritas, realizadas nos últimos anos de forma um pouco desarrumada e por vezes quase casual, talvez possa estar escondido um possível sentido global, ou muito simplesmente a identificação de algum instrumento útil ao filólogo pessoano em particular e, mais genericamente, ao filólogo do manuscrito contemporâneo (ou do que em Itália se chama filologia de autor ou crítica das variantes).

P'ra quê, então? Como já foi sugerido no início do artigo, não para criar um filólogo automático ou virtual, mas para integrar novas ferramentas no trabalho de um filólogo de carne e osso que tente enfrentar a difícil tarefa de reconstruir as fases da escrita da obra de Pessoa.

Confesso, para concluir, que não tenho ainda respostas definitivas. Em compensação, posso afirmar que a resposta já a dei, e pode ler-se nas entrelinhas dos parágrafos anteriores. Reconheço que é difícil ver aqui uma saída airosa. De resto, como escreveu o Mestre Caeiro:

*Eu queria ter o tempo e o socego suficientes
Para não pensar em cousa nenhuma,
Para nem me sentir viver,
Para só saber de mim nos olhos dos outros, reflectido* (Pessoa 2015: 88).

Referências

- Bonsi, Claudia, Eugenio Del Re, Paola Italia, e Michele Ortolani. 2016. «Manuscript & New Technologies. THESMA Project – TeraHERtz & Spectrometry Manuscript Analysis». In *Edizioni critiche digitali/Digital Critical Editions*, ed. Paola Italia, e Claudia Bonsi, 153–160. Roma: Sapienza Università Editrice. http://www.editricesapienza.it/sites/default/files/5369_Italia_Bonsi_EdizioniCriticheDigitali.pdf, 31.07.2018.
- Buscema, Massimo, Riccardo Petritoli, Giovanni Pieri, e Pier Luigi Sacco. 2008. *Auto Contractive Maps*. Roma: Aracne.

- Buscema, Massimo. 2014. «The General Philosophy of Artificial Adaptive Systems (AAS)». In *Archeosema. Artificial Adaptive Systems for the Analysis of Complex Phenomena. Collected Papers in Honour of David Leonard Clarke*, ed. Marco Ramazzotti. *Archeologia e Calcolatori* supplemento 6: 53–84.
- Canettieri, Paolo, Vittorio Loreto, Marta Rovetta, e Giovanna Santini. 2008. «Philology and Information Theory». *Cognitive Philology* 1, <http://www.rspi.uniroma1.it/index.php/cogphil/article/view/8816/8797>, 31.07.2018.
- Canettieri, Paolo. 2012. «Unified theory of the Text (UTT) and the Question of Authorship Attribution». *Memoria di Shakespeare* n.s. 8: 65–77.
- Canettieri, Paolo. 2016. «Le impronte digitali dell'autore. Un metodo di attribuzione automatizzata per i testi delle letterature romanze». *Le forme e la storia* n.s. VI : 229–243.
- Canettieri, Paolo, e Simone Celani. 2018 (no prelo). «Fernando Pessoa e i nuovi metodi attributivi». *Cognitive Philology* 11.
- Castro, Ivo. 2013² (1990¹) *Editar Pessoa*. Lisboa: Imprensa Nacional – Casa da Moeda.
- Celani, Simone. 2004. *La prosa de ficção di Fernando Pessoa: l'esempio de O Caso Vargas. Problemi metodologici e criteri di edizione*. Tese de Doutoramento em Filologia Românica da Universidade da Roma La Sapienza.
- Celani, Simone. 2005. *Il fondo Pessoa. Problemi metodologici e criteri d'edizione*. Viterbo: Sette Città.
- Celani, Simone. 2014. «Artificial Adaptive Systems for Phylological Analysis: the Pessoa Case». In *Archeosema. Artificial Adaptive Systems for the Analysis of Complex Phenomena. Collected Papers in Honour of David Leonard Clarke*, ed. Marco Ramazzotti. *Archeologia e Calcolatori* supplemento 6: 203–215.
- Celani, Simone. 2013. «Quale Pessoa? Ultime edizioni e nuove prospettive». *Critica del Testo* XVI/2: 335–353.
- Celani, Simone. 2015. «I Sistemi Artificiali Adattivi e l'opera di Fernando Pessoa: possibili applicazioni a fini ecdotici e critici». *Semicerchio* LIII/2: 109–116.
- Italia, Paola. 2013. *Editing Novecento*. Salerno: Salerno Editrice.
- Italia, Paola. 2016. «Riscritture gaddiane: da *Eros e la Banda* (1944-45) al *Bugiardione* (1946) a *Eros e Priapo* (1967). In *Riscritture d'autore. La creazione letteraria nelle varianti macro-testuali*, ed. Simone Celani, 7–29. Roma: Sapienza Università Editrice. http://www.editricesapienza.it/sites/default/files/5349_Celani_Riscritture_autore.pdf, 31.07.2018.
- Italia, Paola. 2017. *Come lavorava Gadda*. Roma: Carocci.

- Pessoa, Fernando. 2006. *Il caso Vargas*, ed. Simone Celani. Viterbo: Il Filo.
- Pessoa, Fernando. 2015. *Poemas de Alberto Caeiro*, ed. Ivo Castro. Lisboa: Imprensa Nacional – Casa da Moeda.
- Portela, Manuel, e António Rito Silva. 2016. «Fernando Pessoa's *Book of Disquiet* as a Dynamic Digital Archive». In *Edizioni critiche digitali/Digital Critical Editions*, ed. Paola Italia, e Claudia Bonsi, 23–35. Roma: Sapienza Università Editrice. http://www.editricesapienza.it/sites/default/files/5369_Italia_Bonsi_EdizioniCriticheDigitali.pdf, 31.07.2018.
- Portela, Manuel, e António Rito Silva (eds.). 2017. *Arquivo LdoD: Arquivo Digital Colaborativo do Livro do Desassossego*. Coimbra: Centro de Literatura Portuguesa da Universidade de Coimbra. <https://ldod.uc.pt/>, 31.07.2018.
- Ramazzotti, Marco. 2010. *Archeologia e Semiotica. Linguaggi, codici, logiche e modelli*. Torino: Bollati Boringhieri.
- Ramazzotti, Marco. 2013. «Archeosema. Un modello archeologico per la ricerca teorica, analitica e sperimentale dei fenomeni complessi». *Archeologia e Calcolatori* 24: 283–303.
- Ramazzotti, Marco (ed.). 2014. *Archeosema. Artificial Adaptive Systems for the Analysis of Complex Phenomena. Collected Papers in Honour of David Leonard Clarke*. *Archeologia e Calcolatori* supplemento 6.
- Tomasin, Lorenzo. 2016. «Umanisti scann(erizz)ati». *Il Sole 24 Ore*, 7 luglio 2016, http://www.ilsole24ore.com/art/cultura/2016-07-07/umanisti-scannerizzati-113515.shtml?uuid=ADa6Ivn&refresh_ce=1, 31.07.2018.
- Tomasin, Lorenzo. 2017. *L'impronta digitale. Cultura umanistica e tecnologia*. Roma: Carocci.