Gemma Boleda

ICREA Research Professor
Department of Translation and Language Sciences
Universitat Pompeu Fabra
Roc Boronat 138
08018 Barcelona, Spain

gemma.boleda@upf.edu
+34 93 542 2232
https://gboleda.github.io
ORCID 0000-0001-6140-7080

Education

PhD, Universitat Pompeu Fabra, Spain (*cum laude*)	2007
DEA in Cognitive Science and Language, Universitat Pompeu Fabra, Spain (with Honors)	2003
B.A. in Spanish Philology, Universitat Autònoma de Barcelona, Spain (with Honors)	2000

Professional experience

ICREA Research Professor, Universitat Pompeu Fabra, Spain	2019-
Tenure-track researcher (*Ramón y cajal* fellow), Universitat Pompeu Fabra, Spain	2017-2019
Post-doctoral researcher (*Marie Skłodowska-Curie* fellow), University of Trento, Italy	2015-2016
Post-doctoral researcher and lecturer (*Beatriu de Pinós* fellow), The University of Texas at Austin, US and Universitat Pompeu Fabra, Spain	2012-2015
Researcher, Universitat Pompeu Fabra, Spain	2011-12
Post-doctoral researcher (*Juan de la Cierva* fellow), Universitat Politècnica de Catalunya, Spain	2008-11
Visiting researcher, University of Stuttgart, Germany	2010
Doctoral researcher, Universitat Pompeu Fabra and Barcelona Media Centre d'Innovació, Spain	2001-07
Visiting researcher, Saarland University, Germany	2003,2004
Student assistant, Universitat Pompeu Fabra and Artificial Intelligence Research Institute (IIIA, CSIC), Spain	2000
Student assistant, University of Cologne, Germany	1997-98

Research interests

semantics, concepts, categories, reference, lexicon, semantic theory, language universals, computational semantics, linguistics, cognitive science, computational linguistics, natural language processing, language resources, data science, linguistics, cognitive science, artificial intelligence.

Participation in funded research projects, research groups, and networks

Coordinator, EU, Marie Sklodowska-Curie grant (MSCA European Postdoctoral Fellowships) HORIZON-MSCA-2021-PF-01 101062294 (LIFE - Learning Intentions Following Evolution) Germán Kruszewski, €181,153, Universitat Pompeu Fabra. 2023-2025

Coordinator, Research goup "COLT: Computational Linguistics and 2021-2023

Linguistic Theory" funded by the Catalan government (2021 SGR 00470), €60,000, Universitat Pompeu Fabra.

Collaborating researcher, Canada, NSERC Alliance International - Catalyst Grant ALLRP 576149 (Universal knowledge engines for lexical semantic creativity in humans and machines), 25,000CAD. 13/06/22-12/06/23 — 2022-2023

Co-Principal Investigator, Spain, grant PID2020-112602GB-I00 ("CORE: Contextual effects in the choice of referring expressions for visually presented entities"), €81,070 plus funding for a PhD student. — 2021-2025

Coordinator, EU, Marie Sklodowska-Curie grant H2020-MSCA-IF-2017 790369 (MAGIC - Multimodal Agents Grounded via Interactive Communication), Elia Bruni, €158,121, Universitat Pompeu Fabra. — 2018-2020

Principal Investigator, EU, Horizon 2020 European Research Council Starting Grant 715154 (AMORE: A distributional MOdel of Reference to Entities), €1.5 Million. Universitat Pompeu Fabra. *Success rate: 11% (8% for women).* — 2017-2023

Member, OASIS, CNRS GDRI (International Research Group) Ontology as Structured by the Interfaces with Semantics. PIs: Bridget Copley and Isabelle Roy. — 2017-2020

Coordinator, Research goup COLT: Computational Linguistics and Linguistic Theory funded by the Catalan government (2017 SGR 1575), €19,200, Universitat Pompeu Fabra. — 2017-2019

Host, Horizon 2020 Coordination and Support Action VERTIGO STARTS for Artistic Residency "Blueprints For An Emergent Personality", Kate Aspinall, €15,000, Universitat Pompeu Fabra. — 2017-2018

Member, Ministerio de Economía y Competitividad Redes de Excelencia Program grant FFI2016-81750-REDT (SIGGRAM: Significado y Gramática), €20,000, PI: M. Teresa Espinal, U. Autònoma de Barcelona. — 2017-2018

Principal Investigator, EU Horizon 2020 Marie Curie FI project H2020-MSCA-IF-2014_ST-655577 (LOVe - Linking Objects to Vectors in distributional semantics: A framework to anchor corpus-based meaning representations to the external world), €180,277. University of Trento. — 2015-2016

Principal Investigator, Ministerio de Economía y Competitividad project FFI2014-56672-JIN (Vinculación de objetos con vectores en semántica distributional: Un marco semántico para anclar representaciones del significado basadas en corpus en el mundo externo), €170,000. Universitat Pompeu Fabra, declined. — 2015

Member, Ministerio de Economía y Competitividad Redes de Excelencia Program grant FFI2014-51675-REDT (SIGGRAM: Significado y Gramática), €25.000. PI: Louise McNally, Universitat Pompeu Fabra. — 2015

Collaborating researcher, Ministerio de Economía y Competitividad grant FFI2013-41301-P (CONNECT - Connecting Conceptual and Referential Models of Meaning), €40,000. PI: Louise McNally, Universitat Pompeu Fabra. — 2014-2015

Collaborating researcher, Defense Advanced Research Projects Agency (DARPA) Deep Exploration and Filtering of Text (DEFT) Program grant Statistical Relational Learning and Script Induction for Textual Inference, $1,302,318. PI: Ray Mooney, Co-PI: Katrin Erk, The University of Texas at Austin. — 2012-2014

Collaborating researcher, Ministerio de Ciencia e Innovación grant FFI2010-15006/FILO (OntoSem2 - Natural language ontology and the semantic representation of abstract objects 2), €104,000. PI: Louise McNally, Universitat Pompeu Fabra.	2011-2013
Collaborating researcher, Ministerio de Ciencia e Innovación EXPLORA (Acción Complementaria, modalidad F) grant FFI2010-09464-E (A distributional semantic model for fully recursive phrasal meaning), €32,000. PI: Louise McNally, Universitat Pompeu Fabra.	2011-2012
Member, Research group GLiF: Grup de Lingüística Formal (2014 SGR 698), funded by the Catalan government, €28,500. PI: Josep Quer, Universitat Pompeu Fabra.	2014-2017
Collaborating researcher, Ministerio de Educación y Ciencia grant TIN2009-14715-C04-04 (KNOW2 - Language understanding technologies for multilingual domain-oriented information access), €160,600. PI: Jordi Turmo, Universitat Politècnica de Catalunya.	2010-2012
Member, Research group GPLN: Grup de Processament del Llenguatge Natural"(2009 SGR 1082), funded by the Catalan government. PI: Horacio Rodríguez, Universitat Politècnica de Catalunya.	2009-2013
Member, EU Network of Excellence PASCAL 2: Pattern Analysis, Statistical Modelling, and Computational Learning 2. Scientific Coordinator: John Shawe-Taylor, University College London, UK.	2008-2013
Collaborating researcher, Ministerio de Educación y Ciencia grant HUM2007-60599/FILO (OntoSem - Natural language ontology and the semantic representation of abstract objects), €109,100. PI: Louise McNally, Universitat Pompeu Fabra.	2007-2010
Collaborating researcher, Ministerio de Educación y Ciencia grant TIN2006-15049-C03-03 (KNOW - Developing large-scale multilingual technologies for language understanding), €100,000. PI: Lluís Padró, Universitat Politècnica de Catalunya.	2006-2009
Collaborating researcher, Ministerio de Educación y Ciencia international cooperation grants HA2005-0100 (co-PI: Graham Katz, U. Osnabrück, €11,040) and HF2005-0177 (co-PI: Philip Miller, CNRS-U. Lille, €10,830) ("Natural language ontology for reference to facts and eventualities"). PI: Louise McNally, Universitat Pompeu Fabra.	2006-2007
Member, Research Group Unitat de Recerca en Lingüística (2005 SGR 00067), funded by the Catalan government. PI: Toni Badia, Universitat Pompeu Fabra.	2005-2009
Collaborating researcher, Ministerio de Educación y Ciencia grant HUM2004-05321-C02-02 (ARQUITEXT - Arquitectura integrada para el tratamiento avanzado de textos), €25,920. PI: Toni Badia, Universitat Pompeu Fabra.	2005-2008
Collaborating researcher, EU grant IST–FP6-003768 (METIS-II - Statistical Machine Translation using Monolingual Corpora: from Concept to Implementation), €1,000,000. General Coordinator: Stella Markantonatou (ILSP, Greece); PI at Universitat Pompeu Fabra: Toni Badia.	2004-2007
Collaborating researcher, Ministerio de Ciencia y Tecnología grant TIC2000-1681-C02-01 (PrADo - Sistema de preparación automatizada de documentos), €36,000. PI: Toni Badia, Universitat Pompeu Fabra.	2001-2004

Participation in teaching innovation projects

Collaborating researcher, *Banc de textos traduïts com a recurs didàctic* (BancTrad), 850,000 *pesetas.* PIs: Carme Colominas and Toni Badia, Universitat Pompeu Fabra*.*   2000-02

Student assistant with contract, *Infraestructura informatitzada de recursos lingüístics per a l'ensenyament de la llengua anglesa a la titulació de Traducció i Interpretació*, 1,562,500 *pesetas* in total. PIs: Josep Maria Fontana, Louise McNally, Enric Vallduví, Universitat Pompeu Fabra.   1998-00

Publications

- H-index: 25.[1]

Monographs and edited volumes (marks JCR Q1 journals)*

* **Boleda**; G., A. Herbelot (eds.). 2016. *Special Issue on Formal Distributional Semantics. Computational Linguistics*. *Computational Linguistics* 42:4. *(JCR: Q1, Linguistics, rank 5/180, I. F.: 2.235.)*

Palmer, M., G. **Boleda**, P. Rosso (eds.). 2015. *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*. Denver, Colorado: Association for Computational Linguistics.

Herbelot, A., Zamparelli, R., **Boleda**, G. (eds.) 2013. *Proceedings of the IWCS 2013 Workshop Towards a Formal Distributional Semantics*. Potsdam, Germany: Association for Computational Linguistics.

Artstein, R., G. **Boleda**, F. Keller, S. Schulte im Walde (eds). 2008. *Proceedings of the COLING Workshop on Human Judgements in Computational Linguistics*. Manchester, UK: COLING 2008 Organizing Committee.

Articles in scientific journals (marks JCR Q1 journals)*

* Brochhagen, T., G. **Boleda**. 2022. When do languages use the same word for different meanings? The Goldilocks Principle in colexification. *Cognition*, Volume 226, September 2022, 105179. *(JCR 2020: Q1, rank 18/91 Psychology, Experimental, I.F.: 3.65, JCI: 1.22.)*

Westera, M., A. Gupta, G. **Boleda**, S Padó. 2021. Distributional models of category concepts based on names of category members. *Cognitive Science* 45:9, e13029. *(JCR 2020: Q2, rank 42/90 Psychology, Experimental, I.F.: 2.561)*

* **Boleda**, G. 2020. Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics*, Vol.6: 213-23. DOI: 10.1146/annurev-linguistics-011619-030303. (*JCR: Q1, Linguistics, rank 14/193, I.F.: 3.512.)*

* **Boleda**, G., A. Herbelot. 2016. Introduction to the Special Issue on Formal Distributional Semantics. *Computational Linguistics* 42:4, 619-635. (*JCR: Q1, Linguistics; rank 5/180, I.F.: 2.235.)*

* Corral, A., G. **Boleda**, R. Ferrer i Cancho. 2015. Zipf's Law for Word Frequencies: Word Forms versus Lemmas in Long Texts. PLoS ONE 10(7): doi:10.1371/ journal.pone.0129031. (*JCR: Q1, Science, Multidisciplinary, I.F.: 3.534.)*

* Font-Clos, F., G. **Boleda**, A. Corral. 2013. A scaling law beyond Zipf's law and its relation to Heaps' law. *New Journal of Physics* 15:9, 093033. *(JCR: Q1, Physics, Multidisciplinary, impact factor: 4.063.)*

---

* Boleda, G., S. Schulte im Walde, T. Badia. 2012. Modeling regular polysemy: A study in the semantic classification of Catalan adjectives. *Computational Linguistics* 38(3): 575–616. (*JCR: Q1, Linguistics, rank: 21/166, impact factor: 0.940.*)

Boleda, G., M. Cuadros, C. España-Bonet, M. Melero, L. Padró, M. Quixal, Carlos Rodríguez. 2009. Primera Jornada del Procesamiento Computacional del Catalán. *Revista de Procesamiento del Lenguaje Natural* 43: 387–388.

Boleda, G., S. Schulte im Walde, T. Badia. 2008. An Analysis of Human Judgements on Semantic Classification of Catalan Adjectives. *Research on Language and Computation* 6(3): 247–271.

Boleda, G., T. Badia. 2008. CUCWEB: un corpus de la llengua catalana construït a partir de la web. *Estudis romànics* 30:291-293. Institut d'Estudis Catalans

Mayol, L., G. **Boleda**, T. Badia. 2005. Automatic acquisition of syntactic verb classes with basic resources. *Language Resources and Evaluation* 39(4): 295–312. (*JCR, I. F.: 0.634 in 2009, earliest date I have access to*).

Badia, T., G. **Boleda**, M. Melero, A. Oliver. 2005. El proyecto METIS-II. *Revista de Procesamiento del Lenguaje Natural* 35: 443–444.

Oliver, A., T. Badia, G. **Boleda**, M. Melero. 2005. Traducción automática estadística basada en n-gramas. *Revista de Procesamiento del Lenguaje Natural* 35: 77–84.

McNally, L. and G. **Boleda**. 2004. Relational adjectives as properties of kinds. In Olivier Bonami and Patricia Cabredo Hofherr (eds.) *Empirical Issues in Syntax and Semantics 5*, 179–196.

Alsina, À., T. Badia, G. **Boleda**, S. Bott, À. Gil, M. Quixal, O. Valentín. 2002. CATCG: un sistema de análisis morfosintáctico para el catalán. *Revista de Procesamiento del Lenguaje Natural* 29: 309–310.

Badia, T., G. **Boleda**, Jenny Brumme, C. Colominas, M. Garmendia, M. Quixal. 2002. BancTrad: un banco de corpus anotados con interficie web. *Procesamiento del Lenguaje Natural* 29: 293–294.

***Articles in refereed conference proceedings** (* marks top 10 venues in Computational Linguistics according to Google Scholar;[2] note that 7 out of 10 venues are conferences, because Computational Linguistics / Natural Language Processing is primarily a conference-based field.)*

* Dessì, R., E. Gualdoni, F. Franzon, G. **Boleda**, M. Baroni. 2022. Communication breakdown: On the low mutual intelligibility between human and neural captioning. *Proceedings of EMNLP.*

Gualdoni, E., A. Mädebach, T. Brochhagen, G. **Boleda**. 2022. Woman or tennis player? Visual typicality and lexical frequency affect variation in object naming. *Proceedings of CogSci*, 990-996.

Mädebach, A., E. Gualdoni, E. Torubarova, G. **Boleda**. 2022. Effects of task and visual context on referring expressions using natural scenes. *Proceedings of CogSci*, 3188-3194.

* Aina, L., X. Liao, G. **Boleda**, M. Westera. 2021. Does referent predictability affect the choice of referential form? A computational approach using masked coreference resolution. In *Proceedings of CoNLL 2021*, 454-469.

* Silberer, C., S. Zarrieß, M. Westera, G. **Boleda**. 2020. Humans Meet Models on Object Naming: A New Dataset and Analysis. In *Proceedings of COLING 2020*, 1893-1905.

* Sorodoc, I., K. Gulordava, G. **Boleda**. 2020. Probing for referential information in language models. In *Proceedings of ACL 2020*, to appear.

Westera, M., G. **Boleda**. 2020. A closer look at scalar diversity using contextualized

semantic similarity. In *Proceedings of Sinn und Bedeutung 24*, 439-454.

Aina, L., K. Gulordava, G. **Boleda**. 2020. Modeling word interpretation with deep language models: The interaction between expectations and lexical information. In *Proceedings of CogSci 2020*, 1518-1524.

Gulordava, K., T. Brochhagen, G. **Boleda**. 2020. Deep daxes: Mutual exclusivity arises through both learning biases and pragmatic strategies in neural networks. In *Proceedings of CogSci 2020*, 2089-2095.

* Silberer, C., S. Zarrieß, G. **Boleda**. 2020. Object Naming in Language and Vision: A Survey and a New Dataset. In *Proceedings of LREC 2020*, 5792-5801.

* Aina, L., K. Gulordava, G. **Boleda**. 2019. Putting words in context: LSTM language models and lexical ambiguity. In *Proceedings of ACL 2019 (57th Annual Meeting of the Association for Computational Linguistics)*, 3342-3348, Florence, Italy. Association for Computational Linguistics.

Westera, M., G. **Boleda**. 2019. Don't blame distributional semantics if it can't do entailment. In *Proceedings of IWCS 2019 (Proceedings of the 13th International Conference on Computational Semantics - Long Papers)*, 120-133, Gothenburg, Sweden. Association for Computational Linguistics.

* Del Tredici, M., Fernández, R., **Boleda**, G. 2019. Short-term meaning shift: a distributional exploration. In *Proceedings of NAACL-HLT 2019 (Annual Conference of the North American Chapter of the Association for Computational Linguistics)*, 2069-2075, Minneapolis, USA. Association for Computational Linguistics.

* Aina, L., Silberer, C., Sorodoc, I., Westera, M., **Boleda**, G. 2019. What do entity-centric models learn? Insights from entity linking in multi-party dialogue. In *Proceedings of NAACL-HLT 2019 (Annual Conference of the North American Chapter of the Association for Computational Linguistics)*, 3772-3783, Minneapolis, USA. Association for Computational Linguistics.

* Gulordava, K., L. Aina, and G. **Boleda**. 2018. How to represent a word and predict it, too: improving tied architectures for language modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)*, 2936-2941. Brussels, Belgium. Association for Computational Linguistics.

Boleda, G., S. Padó, N. The Pham, M. Baroni. 2017. Living a discrete life in a continuous world: Reference with distributed representations. In Proceedings of IWCS 2017, Montpellier, France. Association for Computational Linguistics.

* Boleda, G., A. Gupta, S. Padó. 2017. Instances and concepts in distributional space. In *Proceedings of EACL 2017 (15th Conference of the European Chapter of the Association for Computational Linguistics)*, 79-85, Valencia, Spain. Association for Computational Linguistics.

Gupta, A., G. **Boleda**, S. Padó. 2017. Distributed Prediction of Relations for Entities: The Easy, The Difficult, and The Impossible. In *Proceedings of STARSEM 2017*, 104-109, Vancouver, BC. Association for Computational Linguistics.

* Pham, N., G. Kruszewski, G. **Boleda**. 2016. Convolutional Neural Network Language Models. *Proceedings of EMNLP 2016 (Conference on Empirical Methods in Natural Language Processing),* 1153-1162, Austin, US, November. Association for Computational Linguistics.

* Paperno, D., G. Kruszewski, A. Lazaridou, Q. Ngoc, R. Bernardi, S. Pezzelle, M. Baroni, G. **Boleda**, R. Fernandez. 2016. The LAMBADA dataset: Word prediction requiring a broad discourse context. *Proceedings of ACL 2016* (54th Annual Meeting of the Association for Computational Linguistics), 1525-1534, Berlin, Germany, August. East Stroudsburg PA: Association for Computational Linguistics.

* Gupta, A., G. **Boleda**, M. Baroni, S. Padó. 2015. Distributional vectors encode referential

attributes. In *Proceedings of EMNLP 2015 (Conference on Empirical Methods in Natural Language Processing)*, 12-21. East Stroudsburg PA: ACL.

* Roller, S., K. Erk, G. **Boleda**. 2014. Inclusive yet Selective: Supervised Distributional Hypernymy Detection. In *Proceedings of COLING 2014*, Dublin, Ireland, 1025–1036.

Arsenijevic, B., B. Gehrke, G. **Boleda**, L. McNally. 2014. Ethnic adjectives are proper adjectives. In R. Baglini, T. Grinsell, J. Keane, A. R. Singerman, J. Thomas (eds.), *CLS 46-I The Main Session: Proceedings of 46th Annual Meeting of the Chicago Linguistic Society*, Chicago, IL, USA, 17–30.

Beltagy, I., Chau, C., **Boleda**, G., Garrette, D., Erk, K., & Mooney, R. (2013). Montague meets Markov: Deep semantics with probabilistic logical form. In *Proceedings of * SEM*, 11–21.

Boleda, G., M. Baroni, N. The Pham, L. McNally. 2013. Intensionality was only alleged: On adjective-noun composition in distributional semantics. In *Proceedings of IWCS 2013*, Potsdam, Germany, 35–46.

* **Boleda**, G., E. M. Vecchi, M. Cornudella, L. McNally. 2012. First-order vs. higher-order modification in distributional semantics. In *Proceedings of EMNLP-CoNLL 2012*, 1223–1233, Jeju Island, Korea.

* Bruni, E., G. **Boleda**, M. Baroni, N. K. Tran. 2012. Distributional semantics in technicolor. In *Proceedings of ACL 2012*, 136–145, Jeju Island, Korea.

Boleda, G., S. Padó, J. Utt. 2012. Regular polysemy: a distributional model. In *Proceedings of * SEM 2012*, 151–160, Montreal, Canada.

* Melero, M., G. **Boleda**, M. Cuadros, C. España-Bonet, L. Padró, M. Quixal, C. Rodríguez, R. Saurí. 2010. Language technology challenges of a 'small' language (Catalan). In *Proceedings of LREC 2010*, Valletta, Malta.

* Peris, A., M. Taulé, G. **Boleda**, H. Rodríguez. 2010. ADN-classifier: Automatically assigning denotation types to nominalizations. In *Proceedings of LREC 2010*, Valletta, Malta.

* Reese, S., G. **Boleda**, M. Cuadros, L. Padró, G. Rigau. 2010. Wikicorpus: A word-sense disambiguated multilingual Wikipedia corpus. In *Proceedings of LREC 2010*, Valletta, Malta.

* Sánchez-Marco, C., G. **Boleda**, J.M. Fontana, J. Domingo. 2010. Annotation and representation of a diachronic corpus of Spanish. In *Proceedings of LREC 2010*, Valletta, Malta.

* Sanromà, R. and G. **Boleda**. 2010. The Database of Catalan Adjectives. In *Proceedings of LREC 2010*, Valletta, Malta.

* Vandeghinste, V., P. Dirix, I. Schuurman, S. Markantonatou, S. Sofianopoulos, M. Vassiliou, O. Yannoutsou, T. Badia, M. Melero, G. **Boleda**, M. Carl, P. Schmidt. 2008. Evaluation of a Machine Translation System for Low Resource Languages: METIS-II. In *Proceedings of LREC 2008*, Marrakech, Morocco.

* **Boleda**, G., S. Schulte im Walde, T. Badia. 2007. Modelling Polysemy in Adjective Classes by Multi-Label Classification. In *Proceedings of EMNLP-CoNLL 2007*, 171–180, Prague, Czech Republic.

* **Boleda**, G., T. Badia, E. Batlle. 2004. Acquisition of Semantic Classes for Adjectives from Distributional Evidence. In *Proceedings of COLING 2004*, 1119–1125, Geneva, Switzerland.

* Padó, S. and G. **Boleda**. 2004. The Influence of Argument Structure on Semantic Role Assignment. In *Proceedings of EMNLP 2004*, Barcelona, Spain.

* Alsina, À., T. Badia, G. **Boleda**, S. Bott, À. Gil, M. Quixal, O. Valentín. 2002. CATCG: a general purpose parsing tool applied. In *Proceedings of LREC 2002*, Vol. III, 1130–

1135, Las Palmas, Spain.

Articles in refereed edited volumes

McNally, L., G. **Boleda**. 2017. Conceptual vs. Referential Affordance in Concept Composition. In Yoad Winter & James Hampton (eds.) *Compositionality and Concepts in Linguistics and Psychology*, 245-267. Springer. DOI: 10.1007/978-3-319-45977-6.

Baroni, M., G. **Boleda**, S. Padó. 2017. Show me the cup: Reference with continuous representations. To appear in *Lecture Notes in Computer Science* (LNCS) volume, Springer.

Boleda, G., S. Evert, B. Gehrke, L. McNally. 2012. Adjectives as saturators vs. modifiers: Statistical evidence. In M. Aloni, V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, M. Westera (eds.): *Logic, Language and Meaning - 18th Amsterdam Colloquium, Amsterdam, The Netherlands, December 19-21, 2011, Revised Selected Papers*. *Lecture Notes in Computer Science 7218,* 112–121. Springer.

Sánchez Marco, C., G. **Boleda**, J. M. Fontana. 2012. Propuesta de codificación de la información paleográfica y lingüística para textos diacrónicos del español. Uso del estándar TEI. In Torrens Álvarez, M. J. and P. Sánchez-Prieto Borja (eds.), *Nuevas perspectivas para la edición y el estudio de documentos antiguos*, 447–463. Fondo Hispánico de Lingüística y Filología. Vol. 12. Bern (etc.): Peter Lang. ISSN 1663-2648.

Articles in refereed workshop proceedings (marks top 10 venues in Computational Linguistics according to Google Scholar, see fn. 2)*

Sorodoc, I., L. Aina, G. **Boleda**. 2022. Challenges in including extra-linguistic context in pre-trained language models. In *Proceedings of the Third Workshop on Insights from Negative Results in NLP*, 134-138.

Sorodoc, I., G. **Boleda**, M. Baroni. 2021. Controlled tasks for model analysis: Retrieving discrete information from sequences. In *Proceedings of BlackBoxNLP 2021*, 468-478.

Hererra-Palacio, A. M., C. Ventura, C. Silberer, I. Sorodoc, G. **Boleda**, X. Giro-i-Nieto. 2019. Recurrent Instance Segmentation using Sequences of Referring Expressions. In *Proceedings of NeurIPS 2019 Workshops, Visually Grounded Interaction and Language (ViGIL)*. Vancouver, Canada.

* Aina, L., Silberer, C., Sorodoc, I., Westera, M., **Boleda**, G. AMORE-UPF at SemEval-2018 Task 4: BiLSTM with Entity Library. In *Proceedings of SemEval 2018 (12th International Workshop on Semantic Evaluation)*, 65-69, New Orleans, USA. Association for Computational Linguistics. **Winning system in the SemEval-2018 Task 4 competition.**

Sorodoc, I., S. Pezzelle, A. Lazaridou, A. Herbelot, G. **Boleda**, R. Bernardi. 2016. Look, some green circles! Learning to quantify from images. In *Proceedings of the 5th Workshop on Vision and Language at ACL 2016*.

Bernardi, R., G. **Boleda**, R. Fernandez, D. Paperno. 2015. Distributional semantics in use. In *Proceedings of EMNLP 2015 Workshop LSDSem 2015: Linking Models of Lexical, Sentential and Discourse-level Semantics*, 95-101. East Stroudsburg PA: ACL.

Boleda, G., K. Erk. 2015. Distributional Semantic Features as Semantic Primitives – or Not. *Knowledge Representation and Reasoning: Integrating Symbolic and Neural Approaches. Papers from the AAAI Spring Symposium*, 2-5. Palo Alto, California: AAAI Press.

* Beltagy, I., S. Roller, G. **Boleda**, K. Erk, R. Mooney. 2014. UTexas: Natural Language Semantics using Distributional Semantics and Probabilistic Logic. *Proceedings of*

the *International Workshop on Semantic Evaluation SemEval 2014*, 796–801, Dublin, Ireland.

Sánchez-Marco, C., G. **Boleda**, L. Padró. 2011. Extending the tool, or how to annotate historical language varieties. In *Proceedings of the ACL 2011 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2011),* 1–9, Portland, Oregon, USA.

Boleda, G., S. Bott, C. Castillo, R. Meza, T. Badia, V. López. 2006. CUCWeb: a Catalan corpus built from the Web. In *Proceedings of the Second Workshop on the Web as a Corpus at EACL'06,* Trento, Italy.

Badia, T., G. **Boleda**, M. Melero, A. Oliver. 2005. An n-gram approach to exploiting a monolingual corpus for Machine Translation. In *Proceedings of the Second Workshop on Example-based Machine Translation*, *MT Summit X*, Phuket, Thailand.

Boleda, G., T. Badia, S. Schulte im Walde. 2005. Morphology vs. Syntax in Adjective Class Acquisition. In *Proceedings of the ACL-SIGLEX 2005 Workshop on Deep Lexical Acquisition*, Ann Arbor, USA.

Mayol, L., G. **Boleda**, T. Badia. 2005. Automatic learning of syntactic verb classes. In *Proceedings of the Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, 92–97, Saarbrücken, Germany.

Boleda, G., L. Alonso. 2003. Clustering Adjectives for Class Acquisition. In *Proceedings of the 10th EACL Conference Student Research Workshop*, 9–16, Budapest, Hungary.

Badia, T., G. **Boleda**, C. Colominas, M. Garmendia, A. González, M. Quixal. 2002. BancTrad: a web interface for integrated access to parallel annotated corpora. In *Proceedings of the First International Workshop On Language Resources For Translation Work And Research* at LREC 2002, Las Palmas, Spain.

Badia, T., G. **Boleda**, M. Quixal, E. Bofias. 2001. A modular architecture for the processing of free text. In *Proceedings of the Workshop on Modular Programming applied to Natural Language Processing* at EUROLAN 2001, Iasi, Romania, August 2001, 11–18.

Popularization of science

Boleda, G., M. Cuadros, C. España-Bonet, M. Melero, L. Padró, M. Quixal, C. Rodríguez. 2009. El català i les tecnologies de la llengua. *Llengua, Societat i Comunicació* 7: 20–26. ISSN: 1697 5928.

Boleda, G., M. Cuadros, C. España-Bonet, M. Melero, L. Padró, M. Quixal, C. Rodríguez. 2009. Sobre la I Jornada del Processament Computacional del Català. *Llengua i ús* 45: 23–32. ISSN: 1134-7724.

Boleda, G. 2008. Emulant els infants: induint propietats lingüístiques a partir de dades empíriques. *Revista de Catalunya* 235:33–40. ISSN 0213-5876.

Badia, T., G. Boleda, M. Quixal. 2001. Curso sobre Tecnologías de la lengua (segunda edición). *QUARK, Ciencia, Medicina, Comunicación y Cultura* 21:14–16. ISSN: 1135-8521.

Talks

Keynote / invited talks at scientific events

Amsterdam Colloquium. How word meaning and word use shape language. 21/12/22 Keynote.

How word meaning and word use shape language. GLiF seminar, Barcelona, Spain. Invited. 17/11/22

The interplay between cognitive and communicative pressures in the lexicon. 3/08/12

Language, Cognition, and Computation (LCC) group, University of Toronto. Invited.

When do languages use the same word for different meanings? The Goldilocks Principle in the lexicon, Abralin Ao Vivo – Linguists Online. Invited.	12/21
Computational Linguistics and Linguistic Theory, 15th KONVENS (*Konferenz zur Verarbeitung natürlicher Sprache*), Erlangen, Germany. Keynote.	10/19
Generic and situation-specific information in distributed representations, EurNLP conference, London, UK. Invited.	10/19
Talking about you: Deep Learning models of linguistic reference. Workshop on Vector Space Semantics for Dialogue and Discourse, co-located with IWCS 2019, Gothenburg, Sweden. Keynote.	05/19
Distributed methods and the interplay between language and reality. Workshop on meaning relations between phrases and sentences, co-located with IWCS 2019, Gothenburg, Sweden. Keynote.	05/19
Deep learning models of linguistic reference, Multimodal session, Deep Learning Barcelona Symposium, Barcelona. Keynote.	12/18
At the crossroad between discrete and continuous aspects of language, BNAIC/BENELEARN 2018 (Benelux Conference on Artificial Intelligence and Annual Machine Learning Conference of the Benelux), 's-Hertogenbosch, The Netherlands. Keynote.	11/18
AMORE / AI and Language, Frontier Research and Artificial Intelligence conference organized by the ERCEA, Brussels, Belgium. Invited.	10/18
What's in a name? or, what recent results in computational semantics tell us about ambiguity, Ambiguity: Perspectives on Representation and Resolution (ARR) ESSLLI 2018 workshop, Sofia, Bulgaria. Keynote.	08/18
The interplay between conceptual and referential aspects of meaning, Bridging formal and conceptual semantics (BRIDGE-2018) ESSLLI 2018 workshop, Sofia, Bulgaria. Keynote.	08/18
At the crossroad between discrete and continuous aspects of language, VII Jornada Complexitat, Barcelona, Spain. Invited.	05/18
Distributional semantics in linguistic research, Twelfth International Tbilisi Symposium on Language, Logic and Computation (TbiLLC 2017), Lagodekhi, Georgia. Keynote.	09/17
Talking about the world with a distributed model, International Natural Language Generation conference (INLG 2017), Santiago de Compostela, Spain. Keynote.	09/17
Back to our sense, Meeting of the SIGGRAM Network, U. Autònoma de Barcelona, Spain. Invited.	05/17
Back to our sense, 18th Szklarska Poręba Workshop, Poland. Keynote.	03/17
The interplay between conceptual and referential aspects of meaning, NASSLLI 2016 Workshop on Statistical and Logical Models of Meaning (SaLMoM) workshop, New Jersey, US. Keynote.	07/16
Towards linguistic reference with distributional semantics, CREST workshop, Sendai, Japan. Invited.	03/16
From conceptual to referential properties with distributional semantics. IWES:	09/15

International Workshop on Embeddings and Semantics at SEPLN 2015 (Spanish Society for Natural Language Processing conference), Alicante, Spain. Keynote.

Distributional semantics for lexical semantics, Catalonia-Israel Symposium on Lexical Semantics and Grammatical Structure in Event Conceptualization, The Hebrew University of Jerusalem. Invited. 17/02/15

Adquisició de classes semàntiques adjectivals. III Workshop of the PhD Program in Cognitive Science and Language, Barcelona, Spain. Invited. 06/05

Invited talks at scientific institutions

The interplay between cognitive and communicative pressures in the lexicon, at Language, Cognition, and Computation (LCC) group, Department of Computer Science, University of Toronto. 02/08/22

What's in a name? Object naming in humans and machines, CLICS Seminar, CIMeC, University of Trento. 02/05/22

When do languages use the same word for different meanings? The Goldilocks Principle in the lexicon, CogSci Seminar, Queen Mary University London (online). 06/04/22

When do languages use the same word for different meanings? The Goldilocks Principle in the lexicon, Cambridge NLP group, U. Cambridge (online). 02/22

When do languages use the same word for different meanings?, or, the Goldilocks Principle that shapes the lexicon, HHU Colloquium, U. Düsseldorf (online). 01/22

When do languages use the same word for different meanings?, or, the Goldilocks Principle that shapes the lexicon, Research Seminar of the Linguistics Department of the University of Geneva, Switzerland (online). 11/21

Language and use: interaction between referential context and lexicon, Potsdam Computational Linguistics Colloquium, Potsdam University, Potsdam, Germany (online). 06/20

Distributional Semantics and Linguistic Theory, CLASP seminar, University of Gothenburg, Gothenburg, Sweden (online). 04/20

Distributional Semantics and Linguistic Theory, LingLunch seminar series, Department of Linguistics, Université Paris Diderot, Paris, France (online). 03/20

Referential and conceptual aspects of meaning, IMS PhD Retreat, Kleinwalsertal, Austria. 07/19

Invited to participate in the Lorentz Center workshop on Compositionality in Brains and Machines, Leiden, The Netherlands. 08/19

Towards linguistic reference with distributional semantics, INRIA seminar, Lille, France. 02/17

The interplay between sense and reference, Computational Linguistics Seminar, Institute for Logic, Language, and Computation, University of Amsterdam, Amsterdam, The Netherlands. 01/17

Towards linguistic reference with distributional semantics, ILCC/HCRC Seminar Series, Edinburgh, UK. 02/16

Intensionality was only alleged: On adjective-noun composition in distributional semantics. Guest lecture series Sonderforschungsbereich 732: Incremental 07/13

specification in context, Stuttgart University, Germany.

Coloring semantic spaces: Towards perceptually grounded models of word meaning. Guest lecture series Sonderforschungsbereich 732: Incremental specification in context, Stuttgart University, Germany.	02/12
Modeling regular polysemy: A study in the semantic classification of Catalan adjectives. Linguistics Colloquium, University of Texas at Austin, Austin, USA.	09/11
Modeling regular polysemy: A study in the semantic classification of Catalan adjectives. CLIC Research Seminar, CIMeC, Rovereto, Italy.	07/11
Word Sense Disambiguation and regular polysemy. Institutsversammlung, Institut für Maschinelle Sprachverarbeitung, Stuttgart, Germany.	06/10
Computational Feedback to Linguistics: A study in the semantic classification of Catalan adjectives. Nancy NLP Seminar, INRIA-Lorraine, Nancy, France.	03/10
Automatic acquisition of semantic classes for adjectives. Natural Language Processing Seminar, Universitat Politècnica de Catalunya, Barcelona, Spain.	11/07
Acquiring Semantics Classes for Adjectives through Clustering. Computational Linguistics Seminar, King's College, London, UK.	10/04
A Quantitative Approach to the Lexical Semantics of Adjectives. Computational Linguistics Colloquium, Saarland University, Saarbrücken, Germany.	11/04
Acquisition of Semantic Classes for Adjectives. Colloquium of the International Post-Graduate College on Language Technology and Cognitive Systems, Saarland University, Saarbrücken, Germany.	11/04

Refereed conference and workshop presentations
(excluding those published in proceedings or a special volume, see above)

Gualdoni, E., A. Mädebach, T. Brochhagen, G. **Boleda**. 2022. Horse or pony? Visual typicality and lexical frequency affect variability in object naming. Conference of the Society for Computation in Linguistics (SCiL 2022).

Brochhagen, T., G. **Boleda**. 2022. The interaction between cognitive ease and informativeness shapes the lexicons of natural languages. Conference of the Society for Computation in Linguistics (SCiL 2022).

Boleda, G. 2016. Remarks on the CommAI-env. *MAIN@NIPS (MAchine INtelligence workshop, NIPS 2016)*, Barcelona, December 9.

Gupta, A., G. **Boleda**, M. Baroni, S. Pado. 2015. Mapping conceptual features to referential properties. 3rd International ESSENCE Workshop: Algorithms for Processing Meaning, Barcelona (Spain), May 20-22.

Boleda, G., K. Erk. 2014. Distributional semantic features as primitives –or not. NIPS 2014 Workshop on Learning Semantics, Montreal, Canada, December 12.

Boleda, G., S. Evert, B. Gehrke, L. McNally. 2012. A Logistic Regression Model for adjectival modification vs. modification by prepositional phrases. Poster at *Linguistic Evidence*. Tübingen, Germany, February 9-11.

Baroni, M., E. Bruni, G. B. Tran, A. Anderson, G. **Boleda**, M. Ciaramita, A. Lenci. 2011. Multimodal distributional semantics. Poster at *V&L Net Workshop on Vision and Language*. Brighton, UK, September 15.

Berndt, D., G. **Boleda**, B. Gehrke, L. McNally. 2011. Semantic factors in the choice between ethnic adjectives and PP counterparts: Quantitative evidence. Talk at *4th Conference on Quantitative Investigations in Theoretical Linguistics* (QITL-4), Berlin, Germany, March 28-31.

Arsenijevic, B., G. **Boleda**, B. Gehrke, L. McNally. 2010. Unifying the semantics for

"thematic" and "classificatory" uses of ethnic adjectives. Talk at *8èmes Journées Sémantique et Modélisation*, LORIA-INRIA, Nancy, France, March 25-26.

Berndt, D., G. **Boleda**, B. Gehrke, L. McNally. 2009. Nominalizations and nationality expressions: A corpus analysis. Talk at *Corpus Linguistics 2009*, Liverpool, UK, July 20-23.

Boleda, G., Á. Corral, R. Ferrer-i-Cancho, A. Díaz-Guilera. 2009. Word distance distribution in literary texts. Talk at *Corpus Linguistics 2009*. Liverpool, UK, July 20-23.

Boleda, G. 2009. Uso de PLN en otras disciplinas. Talk at *III Jornadas PLN-TIMM: Modelos y técnicas para el acceso a la información multilingüe y multimodal en la web*. Colmenarejo, Spain, February 5-6.

Corral, Á., R. Ferrer-i-Cancho, G. **Boleda**, A. Díaz-Guilera. 2008. Universality classes and community structure in word recurrence. Talk at *BCNet Workshop - trends and perspectives in complex networks*. Barcelona, Spain, December 10-12.

Boleda, G., S. Bott, R. Meza, C. Castillo, T. Badia, V. López. 2005. Usant la web per estudiar el català. Talk at *III Jornades sobre el català a les noves tecnologies*, Barcelona, Spain, April 14-16.

Colominas, C., G. **Boleda**. 2004. The extraction of translationally relevant information from small ad-hoc corpora. Talk at *Third International Conference on Corpus Use and Learning to Translate* (CULT-BCN), Barcelona, Spain, January 22-24.

Alonso, L., G. **Boleda**. 2002. An Approach to Catalan Adjective Lexical Classes by Clustering. Talk at Workshop on Quantitative Investigations for Theoretical Linguistics (QITL), Osnabrück, Germany, October 3-5.

Supervision

Post-doctoral researchers
Andreas Mädebach, 2021-2023.
Thomas Brochhagen, 2019-2022. Moved to tenure-track assistant professor at U. Pompeu Fabra, Spain.
Matthijs Westera, 2017-2020. Moved to tenure-track assistant professor at U. Leiden, The Netherlands.
Elia Bruni, 2018-2020 (Marie Sklodowska-Curie fellow). Moved to substitute professor at U. Osnabrück, Germany, where he became a Full Professor in 2022.
Carina Silberer, 2017-2020. Moved to assistant professor at U. Stuttgart.
Kristina Gulordava, 2018-2019. Moved to research scientist at Shift Technology, France.

PhD theses
Supervisor, Eleonora Gualdoni, 2020-present, Universitat Pompeu Fabra.
Co-supervisor, Xixian Liao, 2019-present, Universitat Pompeu Fabra.
Supervisor, Ionut-Teodor Sorodoc, "Learning to identify and encode entities with deep learning models", 2022, Universitat Pompeu Fabra. Moved to research scientist at Amazon Inc.
Supervisor, Laura Aina, "A Deep Learning Perspective on Linguistic Ambiguity", 2022, Universitat Pompeu Fabra. Moved to research scientist at Amazon Inc.
2nd advisor, Abhijeet Gupta, "Distributional Analysis of Entities", 2021, Stuttgart University. Moved to post-doctoral researcher at Heinrich Heine Universität Düsseldorf, Germany.

Master's theses
Co-supervisor, Marina Bolea, 2022, Universitat Pompeu Fabra.
Co-supervisor, Sophia Harrison, 2022, Universitat Pompeu Fabra.
Co-supervisor, Mar Domínguez, 2022, Universitat Pompeu Fabra.

Co-supervisor, Alberto García, 2020, Universitat Pompeu Fabra.
Co-supervisor, Mar Adrián, 2020, Universitat Pompeu Fabra.
Co-supervisor, Xixian Liao, 2019, Universitat Pompeu Fabra.
Co-supervisor, Jacob N. Adams, 2019, Universitat Pompeu Fabra.
Co-supervisor, Paul-Fabian Kremershof, 2018, Universitat Pompeu Fabra.
Co-supervisor, Anıl Eryılmaz, 2018, Universitat Pompeu Fabra.
Supervisor, Marina Lacambra, 2017, Universitat Pompeu Fabra.
Co-supervisor, Q. Pham Ngoc, 2016, University of Trento.
Co-supervisor, M. Cornudella Gaya, 2012, Universitat Pompeu Fabra.
Supervisor, Samuel Reese, 2009, Institut Supérieur de l'Aéronautique et de l'Espace (ISAE).

Research assistants / visiting PhD researchers:
Shaney Sze, 2023, Barcelona School of Economics, Spain.
Anna Di Natale, 2023, Medical University of Vienna / Graz University of Technology / Complexity Science Hub Vienna, Austria.
Zhirui (Henrietta) Chen, 2022, Peking University, China.
Marco del Tredici, 2017, Amsterdam University, The Netherlands.
Daniel Berndt, 2008-2009, Universitat Politècnica de Catalunya.

Computational linguistics resources (corpora, tools, datasets)

ManyNames, leader. Object naming on a large scale, with 36 human-provided names for each of 25,000 objects in real-world images, freely available at https://github.com/amore-upf/manynames.

LAMBADA, co-leader. Language modeling dataset for word prediction requiring a broad discourse context. 10,000 datapoints, freely available at https://zenodo.org/record/2630551#.XuzTDPJS_5Y.

Wikicorpus, leader. Wikipedia-based trilingual corpus (Catalan, Spanish, English), automatically annotated, over 750 million words, freely available at http://www.cs.upc.edu/~nlp/wikicorpus.

CUCWEB, collaborating researcher. 166-million word Web corpus for Catalan, automatically annotated.

POS-Tagger for Old Spanish, collaborating researcher. Freely available as part of the open source suite of language analyzers FreeLing.

CatCG, collaborating researcher. Constraint Grammar tagger and shallow parser for Catalan.

Five freely available **semantic datasets** on adjective semantics and regular polysemy, leader or co-leader.

Awards, fellowships, and distinctions

Included in the ranking of the 5,000 top female researchers working at Spanish institutions made by the *Consejo Superior de Investigaciones Científicas* (CSIC) - https://www.webometrics.info/en/investigadoras. Rank: 3265.	2022
Leader, winning team at SemEval 2018 Task 4: Character Identification on Multiparty Dialogues	2018
Best Reviewers Recognition, NAACL-HLT 2016	2016

Outstanding Reviewer Recognition, ACL 2015	2015
Ramón y Cajal fellowship, Spain; ranked **1st**, area Philology and Philosophy[3] (success rate: 8.7%)	2015
Marie Skłodowska-Curie Individual Fellowship, EU (success rate: 18.6%)	2015-2017
Marie Curie - Beatriu de Pinós post-doctoral fellowship, EU / AGAUR (success rate: 11.7%)	2012-2015
PASCAL2 European Network of Excellence Internal Visiting Programme fellowship for post-doctoral visit, University of Stuttgart, Germany.	2010
Juan de la Cierva post-doctoral fellowship, MICINN, Spain (success rate: not available)	2008-2011
PhD fellowship, *Fundación Caja Madrid*, Spain (success rate: 14.7%)	2005-2006
Short Research Visit Programme, AGAUR, Spain	2003,2004
PhD fellowship 2001FI 00582, AGAUR, Spain (success rate: 28.3%)	2001-2004
Extraordinary Bachelor Degree Award by the Universitat Autònoma de Barcelona.	2000
Honorable Mention of the National Bachelor Degree Awards by the Spanish Government.	2001
Fellowship for the Introduction to Research of the *Consejo Superior de Investigaciones Científicas* (CSIC; success rate: not available).	2000
Sócrates-Erasmus scholarship, University of Cologne, Cologne, Germany.	1997-1998
Studentship of the Spanish Education Ministry for the first B.A. academic year because of Honours in secondary school.	1994

Organization of scientific events

Chair, AMORE Mini-workshop on Referential Information in Deep Learning Models, Universitat Pompeu Fabra	11/04/2022
Chair, AMORE Mini-workshop on Linguistic Ambiguity and Deep Learning, Universitat Pompeu Fabra	22/03/2022
Co-organizer, GeCKo Symposium (Integrating Generic and Contextual Knowledge), Barcelona	05/2020
Co-organizer, Deep Learning Barcelona Symposium (DLBCN), Barcelona	12/2019
Chair, COLT kick-off workshop, Barcelona	12/2018
Co-organizer and chair of Language session, Deep Learning Barcelona Symposium (DLBCN), Barcelona.	12/2018
Co-organiser, IWCS 2013 Workshop: Towards a formal distributional semantics, Potsdam, Germany.	2013
Co-organiser, Workshop on Computational Processing of Catalan (*Jornada del Processament Computacional del Català*), Barcelona.	2009
Co-organiser, Nanoworkshop on statistical physics and linguistics, Barcelona, March.	2009

---

3   https://sede.micinn.gob.es/stfls/eSede/Ficheros/2016/
     RESOLUCION_ELEGIBLES_Y_RESERVAS_RYC_2015_31_08_2016.pdf, Spanish name of area:
     "Filología y Filosofía".

Co-organiser, *COLING 2008 workshop on human judgements in Computational Linguistics*, Manchester, UK.	2008
Organiser, reading group on Computational Semantics at Universitat Politècnica de Catalunya.	2008-11
Organiser, GLiCom's seminar on Computational Linguistics at Universitat Pompeu Fabra.	2002-04

Dissemination and outreach

Mentor, Amsterdam Colloquium 2022 Pop-Up Mentoring Program (PUMP)	2022
Mentor, CogSci 2022 Mentoring Program	2022
Co-organizer, *Jornada STEM: visibilitzem l'enginyeria* (STEM day: visibilizing Engineering)	2022
Mentor, Mentoring Sessions of Deep Learning Barcelona Symposium (DLBCN 2021)	2021
Mentor, Mentoring Program of WiNLP 2019 workshop (Widening NLP).	2019
Mentor, MENTOS Mentoring Female Engineering Students Program of UPF's Technology Department.	2018

Service / other professional activities

Area Chair, ACL 2023	2023
Standing review committee member, *Transactions of the Association for Computational Linguistics (TACL)* journal.	2017-2022
Area Chair for Lexicon & Semantics, COLING 2022	2022
Collaborator in activites related to fostering ERC grants in Spain, organized by the Spanish Science and Technology Foundation (*Fundación Española para la Ciencia y la Tecnología* – FECYT): • Talk @ERC Reading Days to share know-how, Sept 17 2021 • Adviser for ERC proposals (StG, AdG) • Jury member for Role-play interview (AdG)	2021, 2022
Jury member for Role-play interviews for ERC Starting Grant, AGAUR, Barcelona.	2017
Area Chair for Semantics, ACL 2016, Berlin, Germany	2016
Co-organiser, ESSLLI 2016 workshop *DSALT: Distributional Semantics and Linguistic Theory* (Bolzano, Italy, 15-19 August 2016).	2016
Co-editor, Special Issue on Formal distributional semantics, *Computational Linguistics* journal, MIT Press, USA.	2015
Program co-chair, *Fourth Joint Conference on Lexical and Computational Semantics* (* SEM 2015), Denver, Colorado, June 4-5 2015.	2015
Local co-chair, *27th European Summer School in Logic, Language and Information* (ESSLLI 2015), Barcelona, Spain, August 3-14 2015.	2015
Information Officer, SIGSEM Board (Special Interest Group on Computational Semantics), Association for Computational Linguistics.	2013-2020
Area Chair, *Second Joint Conference on Lexical and Computational*	2013

Semantics (* SEM 2013), Atlanta, USA, June 13-14 2013.

Evaluation

PhD thesis committees:
- Timothee Mickus, *On the Status of Word Embeddings as Implementations of the Distributional Hypothesis*, Université de Lorraine, France, March 2022.
- Fei Liu, *Memory-augmented neural networks for better discourse understanding*, The University of Melbourne, Australia, February 2020.
- Pranava S. Madhyastha, *Exploiting word embeddings for modeling bilexical relations*, PhD thesis, Universitat Politècnica de Catalunya, July 2017.
- Lauren Romeo, *The structure of the lexicon in the automatic acquisition of lexical information*, PhD thesis, Universitat Pompeu Fabra, 2015.
(Plus 2-3 MA theses committees per year since 2017)

Project evaluation: Dutch Research Council (NWO) 2021; ERC Starting Grant 2020; Spanish Research Agency (*Agencia Estatal de Investigación; Programa Estatal de Generación del Conocimiento y de Retos*) 2019; Research Agency of Argentina (*Fondo para la Investigación Científica y Tecnológica* - FONCYT) 2015; Polish National Science Centre (*Narodowe Centrum Nauki* - NCN), 2015.

Position evaluation: Postdoctoral Researcher Fellow in Computational Linguistics, PAR 2019/244, University of Gothenburg (2019); Junior Research Professor, University of Leuven (2019), Postdoctoral Researcher in Linguistics, University of the Basque Country (2017).

Reviewing

Journals and books: *Open Research Europe, Applied Linguistics, Linguistics**, ACL Rolling Review, Dialogue and Discourse, Cognitive Science, Corpus Linguistics and Linguistic Theory, Nature Human Behavior, SKASE Journal of Theoretical Linguistics, Transactions of the ACL (TACL), ACM Computing Surveys, Artificial Intelligence, Semantics and Pragmatics, Natural Language Engineering, ACM Transactions on Speech and Language Processing, Computational Linguistics, Language Resources and Evaluation, Corpora*, Springer, John Benjamins.

Conferences: Amsterdam Colloquium 2022, CogSci 2022, LSA 2021, LSA 2020, EMNLP 2020, *SEM 2020, COLING 2020, *Sinn und Bedeutung* 2020, ConSOLE 28, *Sinn und Bedeutung* 2018, *SEM 2017, EMNLP 2017, ACL 2017, EACL 2017, IWCS 2017, LaML, LACL 2016, NAACL-HLT 2016 (San Diego, USA), EMNLP 2016 (Austin, USA), ACL-IJCNLP 2015 (Beijing, China), IWCS 2015 (London, UK), EMNLP 2015 (Lisbon, Portugal), ACL 2014 (Baltimore, USA), EACL 2014 (Gothenburg, Sweden), COLING 2014 (Dublin, Ireland), *SEM 2014 (Dublin, Ireland), ACL 2013 (Sofia, Bulgaria), IWCS 2013 (Potsdam, Germany), GL2013 (Pisa, Italy), *SEM 2012 (Montreal, Canada), LREC 2012 (Istanbul, Turkey), EACL 2012 (Avignon, France), IJCNLP 2011 (Chiang Mai, Thailand), ACL-HLT 2011 (Portland, Oregon, USA), COLING 2010 (Beijing, China), LREC 2010 (Valletta, Malta), EMNLP 2009 (Singapore), SEPLN 2009 (Donostia, Spain), NODALIDA 2009 (Odense, Denmark), EACL 2009 (Athens, Greece), ACL 2008 (Columbus, Ohio, USA).

Workshops: NAACL Student Research Workshop 2022, End-to-End Compositional Models of Vector-Based Semantics workshop at ESSLLI 2022, Bridges and Gaps between Formal and Computational Linguistics workshop at ESSLLI 2022 (BriGap), BlackBoxNLP 2021, DeeLIO at EMNLP 2020, RepEval 2019, Workshop on Continuous Vector Space Models and their Compositionality at EACL 2014 (Gothenburg, Sweden),

EACL 2014 Student Research Workshop (Gothenburg, Sweden), ACL 2013 Student Research Workshop (Sofia, Bulgaria), The 1st Workshop on Metaphor in NLP at NAACL 2013 (Atlanta, USA), EACL 2012 Student Research Workshop (Avignon, France), ESSLLI 2011 Student Session (Ljubljana, Slovenia), First Workshop on Computational Neurolinguistics at NAACL-HLT 2010 (Los Angeles, USA), ESSLLI 2010 Compositionality and Distributional Semantic Models Workshop (Copenhagen, Denmark), CBA 2010 (Corpus-Based Approaches to Paraphrasing and Nominalization 2010; Barcelona, Spain), ESSLLI 2008 Distributional Lexical Semantics Workshop (Hamburg, Germany), ACL 2007 Student Research Workshop (Prague, Czech Republic), Workshop on Contextual Information in Semantic Space Models (CoSMo 2007): Beyond Words and Documents (Roskilde University, Denmark), EACL 2006 Student Research Workshop (Trento, Italy), Penn Linguistics Colloquium (2005, 2006).

Summer schools: *Escola d'Estiu de Lingüística Catalana* 2021, 2022 (Summer School in Catalan Linguistics), NASSLLI 2018.

Teaching

University of Texas at Austin

Syntax and Semantics: The Structure and Meaning of Utterances. B. A. Linguistics, Spring.	2014

Universitat Pompeu Fabra
(Role: up to 2006-07 Teaching Assistant, all other courses lecturer.)

Mètodes Empírics en l'Estudi del Llenguatge / Empirical Methods for the Study of Language. B.A. Applied Languages.	2017-19
Computational Semantics. Master's *Lingüística Teòrica i Aplicada.*	2016-23
Anivellament / Levelling 1. Master's *Lingüística Teòrica i Aplicada.*	2014-15
Estructura de les Llengües / Language Structure. B.A. Applied Languages.	2014-15
Pragmàtica i Semàntica Computacional / Computational Semantics and Pragmatics. B.A. Linguistics and Technological Applications and Master's Translation studies.	2010-11, 2008-09
Introducció a la Lingüística Computacional: Aprendre a programar en Prolog / Introduction to Computational Linguistics: Learning to program in Prolog. B.A. Linguistics and Technological Applications and Master's Translation studies.	2009-10, 2006-07
Informàtica Aplicada a la Traducció / Informatics Applied to Translation. B.A. Translation and Interpreting.	2003-04
Noves Tecnologies i Traducció / New Technologies and Translation. B.A. Translation and Interpreting and Master's Translation studies.	2006-07, 2002-03
Sistemes de Traducció Automàtica / Machine Translation Systems. B.A. Translation and Interpreting.	2001-03
Lingüística Computacional II / Computational Linguistics II. B.A. Translation and Interpreting.	2001-02
Lingüística Computacional I / Computational Linguistics I. B.A. Translation and Interpreting.	2001-02

Other courses

Tutorial on Computational Linguistics at the CONSOLE 28 conference (co-	01/2019

taught with Matthijs Westera and Laura Aina), Barcelona, Spain, January 28

Course on Computational Lexical Semantics at 21st European Summer School   07/2009
in Logic, Language and Information (ESSLLI 2009), Bordeaux, France, July
20-31 (co-taught with Stephanie Evert).

Other merits

Acreditació de recerca, Agència per a la qualitat del sistema universitari de   2012
Catalunya (AQU).

Membership in scientific associations

Cognitive Science Society	2020-pres.
Asociación de Mujeres Investigadoras y Tecnólogas (AMIT-Cat)	2017-pres.
Association for Computational Linguistics (ACL)	2004-pres.
Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)	1999-2000, 2009-pres.

Languages

Catalan (native), Spanish (native), English (proficient), German (excellent command),
French (good command), Italian (medium command).

Last modified: February 2023