

Marco Baroni

Facebook Artificial Intelligence Research
6 Rue Ménars
75002 Paris
France

mbaroni@gmail.com
<http://clic.cimec.unitn.it/marco>
<https://scholar.google.com/citations?user=l-xu2w0AAAAJ&hl=en>

Google Scholar statistics

- Total citations: 8179 (5878 since 2013)
- H-index: 43 (35 since 2013)

Education

- Ph.D. in Linguistics, University of California, Los Angeles, 2000
- M.A. in Linguistics, University of California, Los Angeles, 1997
- B.A. cum laude in Linguistics, University of Padua, 1995

Professional experience

- Research scientist and research manager, Facebook Artificial Intelligence Research (FAIR), Paris (France), 2016-present
- Associate professor (tenured researcher until February 2013), Center for Mind/Brain Sciences (CIMEC), University of Trento (Italy), 2006-present (on leave)
- Tenured researcher, Dipartimento di Studi Interdisciplinari su Traduzione, Lingue e Cultura (SITLEC), University of Bologna (Italy), 2002-2006
- Researcher, Austrian Research Institute for Artificial Intelligence (ÖFAI), Vienna (Austria), 2001-2002
- Computational linguist, Conversay, Redmond (USA), 2000-2001

Service

- Editorial board of the Transactions of the Association for Computational Linguistics (TACL), 2017-2019 (in standing committee of reviewers 2012-2017)
- Editorial board of Computational Linguistics journal, 2014-2016
 - TACL and Computational Linguistics are the two top journals in computational linguistics (Google Scholar venue ranking)
- Program chair of *SEM 2013
- Area chair of EMNLP 2015, EACL 2014, CLIC.it 2014, *SEM 2012
- Elected information officer of SIGSEM, the Special Interest Group on Semantics of the Association for Computational Linguistics, 2013-2019
- Elected secretary of SIGWAC, Special Interest Group on the Web as Corpus of the Association for Computational Linguistics, 2007-2009

- Reviewer of 22 research projects, including assignments by national science councils of Israel, the Netherlands, Italy, Canada, UK, Belgium, US and by the ERC (starting and advanced grants)
- In program committee of 40 conferences and 38 workshops (best reviewer award at EMNLP 2010)
- Reviewer of more than 30 journal articles, 1 book and 3 articles in edited volumes

Funded projects and fellowships

- LOVE, Marie Sklodowska-Curie project, 2015-2017, supervisor
- COMPOSES, ERC Starting Grant, 2011-2016, PI
- European Network on Integrating Vision and Language (iV&L Net), ICT COST Action, 2014-2018, Management Committee member and dissemination officer until 2015
- Web 2.0 as Corpus, Google Research Award, 2010-2013, PI
- Conceptual Representation in the Blind: Empirical Data and Computational Simulations, national Italian PRIN project, 2010-2012, PI
- PAISA', national Italian FIRB project, 2009-2011, PI
- LiveMemories, Trento Province PAT project, 2008-2010, participant
- LiMiNE, University of Bologna strategic program, 2007-2009, external consultant
- Invited Scholar Fellowship, the National Institute for Japanese Language, Tokyo, 2007
- CompoNet, national Italian PRIN project, 2006-2008, participant
- Marco Polo Scholarship for a research period abroad, University of Bologna, 2005
- Chancellor Fellowship, University of California, Los Angeles, 1995-2000
- Education Abroad Program Fellowship, University of California, Los Angeles, 1993-1994

Workshop/shared task organization

- Machine Intelligence Workshop at NIPS, Barcelona (2016)
- SemEval Task 1 (Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment) (2014)
- ESSLLI Distributional Lexical Semantics Workshop, Hamburg (2008)
- Contextual Information in Semantic Space Models Workshop at Context 07, Roskilde, (2007)
- CLEAN EVAL shared task on automated cleaning of Web data (2007)
- The Web as Corpus Workshop at EACL, Trento (2006)
- Workshop on The Web as Corpus at Corpus Linguistics, Birmingham (2005)

Selected publicly available tools and resources

See http://clic.cimec.unitn.it/marco/tools_and_resources.html for full list and links.

- CommAI-env: interactive environment to train/test communication-based AIs

- SCAN: a language-driven navigation environment to evaluate compositional learning and zero-shot generalization
- WaCky: large (more than one billion words), linguistically annotated Web corpora in English, French, German and Italian
- DISSECT: toolkit to construct and compose distributed semantic representations
- LAMBADA: data set for testing broad context understanding of language models
- SICK: data set for large-scale evaluation of compositional semantic models
- MEN: evaluation set for multimodal lexical semantic models
- BLIND: semantic norms collected from congenitally blind and sighted subjects
- BLESS: data set to evaluate computational models of semantic similarity
- zipfR: toolkit for lexical statistics in R
- BootCaT: toolkit for bootstrapping corpora from the web
- Morph-it!: Italian morphological lexicon
- La Repubblica-SSLMIT: a large annotated corpus of Italian newspaper text

Keynotes and selected invited lectures

See http://clic.cimec.unitn.it/marco/baroni_short_cv.html for full list.

- M. Baroni. 2018. Artificial neural networks and the challenge of compositional generalization. Invited lecture at the Collège de France, Paris (France) and at the Institute for Logic, Language and Computation, University of Amsterdam (the Netherlands).
- M. Baroni. 2017. Spectacular successes and failures of recurrent neural networks applied to language. Keynote at the Fourth Italian Conference on Computational Linguistics (CLIC-it), Rome (Italy) and at the Paris Syntax and Semantics Conference.
- M. Baroni. 2017. Learning to generalize by skill composition. Keynote at the First Conference on Logic and Machine Learning in Natural Language (LaML), Gothenburg (Sweden).
- M. Baroni. 2017. Simple tasks, grand challenges: defining an evaluation roadmap for general AI. Keynote at the Third Research and Applied AI Summit, London (UK), at CICLing 2017, Budapest (Hungary), and invited lecture at the TiCC Colloquium, Tilburg University (the Netherlands), in the Synthetic Language Learner seminar series, LSCP-ENS, Paris (France), at the From Computational Modelling to Behavior via Multimodal Corpus Data symposium of ICPS 2017, Vienna (Austria), in the Laboratoire Lattice seminar series, CNRS, Paris (France) and in the Language Technology Lab Seminar series, University of Cambridge (UK).
- M. Baroni. 2017. Linguists on the verge of a nervous breakdown. Invited special event talk at CICLing 2017, Budapest (Hungary).
- M. Baroni. 2016. End-to-end conversational agents: what's missing? Invited talk at the Let's Discuss Workshop of NIPS 2016, Barcelona (Spain).
- M. Baroni. 2016. Will computers ever be able to chat with us? Evening lecture at ESSLLI 2016, Bolzano (Italy).
- M. Baroni. 2016. Living a discrete life in a continuous world. Invited talk at the ESSLLI 2016 DSALT workshop, Bolzano (Italy).
- M. Baroni. 2016. The role of simulation in training end-to-end conversational agents. Invited talk at the Deep Learning Workshop at ICML16, New York (USA).
- M. Baroni. 2015. Computational systems that learn linguistic meaning composition from natural data. Invited talk at the Bridging Neural Mechanisms and Cognition FENS Spring Brain Conference, Rungstedgaard, Copenhagen (Denmark).

- M. Baroni. 2015. Grounding distributional semantics in the visual world. Invited talk at the EMNLP Workshop on Vision and Language (VL15), Lisbon (Portugal).
- M. Baroni. 2015. Grounding word representations in the visual world. Invited talk at the ERC Allegro Workshop, INRIA, Grenoble (France).
- M. Baroni. 2015. What (multimodal) distributional semantic models learn during their childhood. Joint keynote at *SEM 2015 and SemEval-2015, Denver (Colorado, USA).
- M. Baroni. 2015. The grumpy linguist's guide to distributed representations of sentence meaning. Invited talk at the NAACL 2015 Workshop on Vector Space Modeling for NLP, Denver (Colorado, USA) and at the Meaning in Context Symposium (CAS-LMU, Munich, Germany).
- M. Baroni. 2015. Vector-based semantics from joint linguistic and visual evidence. Invited talk at the Interaction between Formal and Distributional Semantics Workshop, IRIT, Toulouse (France).
- M. Baroni. 2014. Linking vectors to the world: Multimodal and cross-modal distributional semantics. Invited lecture in the TLA e-Humanities in Action lecture series, Max Planck Institute for Psycholinguistics, Nijmegen (Holland).
- M. Baroni. 2014. DEcompositional distributional semantics. Invited lecture in the CIS seminar series, Ludwig-Maximilians-Universität, Munich (Germany) and in the GLIF seminar series, Universitat Pompeu Fabra, Barcelona (Spain).
- M. Baroni. 2014. Multimodal and cross-modal distributional semantics: Towards a common semantic space for words and things. Keynote at Dialogue 2014, Bekasovo (Russia).
- M. Baroni. 2014. Composition in distributional semantics. Invited lecture in the ABBYY Open Seminars series, ABBYY Headquarters, Moscow (Russia).
- M. Baroni. 2014. Learning meaning representations from naturally occurring text and images. Invited talk at the SLI Usage-based Theories and Approaches in Linguistics Conference, Bolzano (Italy).
- M. Baroni. 2014. Your first encounter with a wampimuk: Cross-modal mapping of words and things. Invited talk at GET 2014, the first Google Event in Trento, DISI, Trento (Italy).
- M. Baroni. 2014. Bringing distributional semantics out in the (visual) world. Invited talk at the Department of Linguistics of the University of Potsdam (Germany).
- M. Baroni. 2013. My love affair with the Web, and why it ended. Invited talk at BOTWU, Forlì (Italy).
- M. Baroni. 2012. Compositionality in (high-dimensional) space. Keynote at KONVENS 2012, Vienna (Austria).
- M. Baroni. 2012. Compositional operations to represent phrases and sentences in distributional semantics. Invited talk at LSD 2012 (Leuven Statistics Days), KU Leuven (Belgium).
- M. Baroni. 2012. Distributional semantics with eyes: Enriching corpus-based models of word meaning with automatically extracted visual features. Invited talk at the Computational Linguistics Invited Colloquium of the Department of Computational Linguistics and Phonetics of Saarland University, Saarbrücken (Germany).
- M. Baroni. 2012. You can tell a word by the (visual) company it keeps: Extracting lexico-semantic information from text and images. Invited talk at the Internet Lexicography workshop, EURAC, Bolzano (Italy).
- M. Baroni. 2012. Making corpus-based semantics more human-like. Invited talk in the Cognition, Perceptual and Brain Sciences seminar series, UCL, London (UK).
- M. Baroni. 2012. Distributional semantics beyond and above words. Keynote at Linguistic Evidence 2012, Tübingen University (Germany).
- M. Baroni. 2012. Distributional semantics in the phrasal and sentential domains: How and why. Keynote at CLIN 2012, Tilburg University (the Netherlands).

- M. Baroni. 2012. You shall know a word by the (visual) company it keeps: Towards a multimodal distributional semantics. Invited seminar in the CLiPS Colloquia series, University of Antwerp (Belgium).
- M. Baroni. 2010. The PAISA' project. Invited talk at the Human Language Technologies Workshop of AI*IA 2010, Brescia (Italy).
- M. Baroni. 2010. Web 2.0 as corpus: One decade of textual analysis with Web data. Keynote at JADT 2010, Sapienza University, Rome (Italy).
- M. Baroni. 2008. Distributional semantics: From ad hoc solutions to persistent models. Keynote at IS-LTC 2008, Stefan Institute, Ljubljana, Slovenia.
- M. Baroni. 2008. Distributional semantics. Invited talk at the Beyond Short Units workshop of the "Sound to Sense" Marie Curie Research Training Network, Federico II University, Naples, Italy.
- M. Baroni. 2008. From word co-occurrences to properties of concepts: Using corpora to simulate the human experience. Invited talk at the Corpus Tools in Teaching and Research Colloquium, Bressanone/Brixen, Italy.
- M. Baroni. 2007. Extracting structured semantic spaces from corpora. Invited talk at the National Institute for Japanese Language, Tokyo, Japan.
- M. Baroni. 2007. Building large linguistic corpora by Web crawling. Invited lecture at the Dipartimento di Scienze Letterarie e Filologiche, University of Turin, Italy.
- M. Baroni. 2006. Building large corpora from the Web. Invited talk in the Language and Communication Technologies colloquium series, University of Bolzano, Italy.
- M. Baroni. 2005. Spotting translated text: Humans vs. (support vectors) machines. Invited talk at the Cognitive Science Institute, University of Osnabrück, Germany.
- M. Baroni. 2004. Using the web as a source of linguistic data: Experiences, problems and perspectives. Invited lecture at the Humboldt University, Berlin, Germany.
- M. Baroni. 2003. Metodi non-supervisionati per la scoperta di morfemi e relazioni morfologiche. Invited lecture at the Istituto di Linguistica Computazionale, CNR, Pisa, Italy.

Publications

See <http://clic.cimec.unitn.it/marco/research.html> for complete list and full text of most articles. Publications with more than 100 citations in Google Scholar are marked with +++, except in section 1), where the exact citation count is reported.

1) Top 5 most cited publications

These entries are also repeated below in the appropriate sections, with further bibliometric details.

- M. Baroni, S. Bernardini, A. Ferraresi and E. Zanchetta. 2009. The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora. *Journal of Language Resources and Evaluation* 43(3). **832** citations.
- M. Baroni, G. Dinu and G. Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors *Proceedings of ACL*. **645** citations.
- M. Baroni and A. Lenci. 2010. Distributional Memory: A general framework for corpus-based semantics. *Computational Linguistics* 36(4). **487** citations.

- M. Baroni and S. Bernardini. 2004. BootCaT: Bootstrapping corpora and terms from the web. Proceedings of LREC. **433** citations.
- M. Baroni and R. Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. Proceedings of EMNLP. **345** citations.

2) *Journal articles*

Most recent Scopus journal rank reported after each entry, if available.

- A. Lazaridou, M. Marelli and M. Baroni. 2017. Multimodal word meaning induction from minimal exposure to natural text. *Cognitive Science*. 41(S4). (Experimental and Cognitive Psychology #21/121)
- E. Vecchi, M. Marelli, R. Zamparelli and M. Baroni. 2017. Spicy adjectives and nominal donkeys: Capturing semantic deviance using compositionality in distributional spaces. *Cognitive Science* 41(1). (Experimental and Cognitive Psychology #21/121)
- J. Hernandez-Orallo, M. Baroni, J. Bieger, N. Chamit, D. Dowe, K. Hofmann, F. Martinez-Plumed, C. Strannegard and K. Thorisson. 2017. A new AI evaluation cosmos: Ready to play the game? *AI Magazine* 38(3). (Artificial Intelligence #65/152)
- G. Kruszewski, D. Paperno, R. Bernardi and M. Baroni. 2016. There is no logical negation here, but there are alternatives: modeling conversational negation with distributional semantics. *Computational Linguistics* 42(4). (Language and Linguistics #6/582)
- D. Paperno and M. Baroni. 2016. When the whole is less than the sum of its parts: How composition affects PMI values in distributional semantic vectors. *Computational Linguistics* 42(2). (Language and Linguistics #6/582)
- M. Baroni. 2016. Grounding distributional semantics in the visual world. *Language and Linguistics Compass* 10(1). Invited article.
- P. Tremblay, I. Deschamps, M. Baroni and U. Hasson. 2016. Neural sensitivity to syllable frequency and mutual information in speech perception and production. *NeuroImage* 136. (Neurology #4/138)
- L. Bentivogli, R. Bernardi, M. Marelli, S. Menini, M. Baroni and R. Zamparelli. 2016. SICK through the SemEval glasses: Lessons learned from the evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment. *Journal of Language Resources and Evaluation* 50(1). (Language and Linguistics #35/582)
- M. Marelli and M. Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122(3). (Cognitive Psychology #5/181)
- A. Anderson, E. Bruni, A. Lopopolo, M. Poesio and M. Baroni. 2015. Reading visually embodied meaning from the brain: visually grounded computational models decode visual-object mental imagery induced by written text. *NeuroImage* 120. (Neurology #4/138)
- G. Kruszewski, D. Paperno and M. Baroni. 2015. Deriving Boolean structures from distributional vectors. *Transactions of the Association for Computational Linguistics* 3.
- A. Lazaridou, G. Dinu, A. Liska and M. Baroni. 2015. From visual attributes to adjectives through decompositional distributional semantics. *Transactions of the Association for Computational Linguistics* 3.
- F.M. Zanzotto, L. Ferrone and M. Baroni. 2015. When the whole is not greater than the combination of its parts: A decompositional look at compositional distributional semantics. *Computational Linguistics* 41(1). (Language and Linguistics #6/582)

- M. Marelli, G. Dinu, R. Zamparelli and M. Baroni. 2015. Picking buttercups and eating buttercups: Spelling alternations, semantic relatedness and their consequences for compound processing. *Applied Psycholinguistics* 36(6). (Language and Linguistics #46/582)
- D. Paperno, M. Marelli, K. Tentori and M. Baroni. 2014. Corpus-based estimates of word association predict biases in judgment of word co-occurrence likelihood. *Cognitive Psychology* 74. (Experimental and Cognitive Psychology #5/121)
- M. Baroni, R. Bernardi and R. Zamparelli. 2014. Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies* 9(6). +++
- E. Bruni, N. Tran and M. Baroni. 2014. Multimodal distributional semantics. *Journal of Artificial Intelligence Research* 49. **2017 IJCAI-JAIR best paper prize**. (Artificial Intelligence #17/152) +++
- M. Hernandez, S. Fairhall, A. Lenci, M. Baroni and A. Caramazza. 2014. Predication drives verb cortical signatures. *Journal of Cognitive Neuroscience* 26(8). (Cognitive Neuroscience #20/84)
- A. Lenci, M. Baroni, G. Cazzolli and G. Marotta. 2013. BLIND: a set of semantic feature norms from the congenitally blind. *Behavior Research Methods* 45(4). (Experimental and Cognitive Psychology #16/121)
- M. Baroni. 2013. Composition in distributional semantics. *Language and Linguistics Compass* 7(10). Invited article.
- B. Magnini, M. Baroni, M. Federico and R. Navigli. 2013. Recent advancements in human language technology in Italy. *Intelligenza Artificiale*, VII-2.
- P. Tremblay, M. Baroni and U. Hasson. 2013. Processing of speech and non-speech sounds in the supratemporal plane: Auditory input preference does not predict sensitivity to statistical structure. *NeuroImage* 66. (Neurology #4/138)
- A. Herdagdelen and M. Baroni. 2012. Bootstrapping a game with a purpose for commonsense collection. *ACM Transactions on Intelligent Systems and Technology* 3(4). (Theoretical Computer Science #7/113)
- A. Herdagdelen and M. Baroni. 2011. Stereotypical gender actions can be extracted from Web text. *Journal of the American Society for Information Science and Technology* 62(9). (Library and Information Sciences #15/194)
- G. Kremer and M. Baroni. 2011. A set of semantic norms for German and Italian. *Behavior Research Methods* 43(1). (Experimental and Cognitive Psychology #16/121)
- M. Baroni and A. Lenci. 2010. Distributional Memory: A general framework for corpus-based semantics. *Computational Linguistics* 36(4). (Language and Linguistics #6/582) +++
- M. Baroni, B. Murphy, E. Barbu and M. Poesio. 2010. Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science* 34(2). (Experimental and Cognitive Psychology #21/121) +++
- M. Baroni, S. Bernardini, A. Ferraresi and E. Zanchetta. 2009. The WaCky Wide Web: A collection of very large linguistically processed Web-crawled corpora. *Journal of Language Resources and Evaluation* 43(3). (Language and Linguistics #35/582) +++
- M. Baroni, E. Guevara and R. Zamparelli. 2009. The dual nature of deverbal nominal constructions: Evidence from acceptability ratings and corpus analysis. *Corpus Linguistics and Linguistic Theory* 5(1). (Language and Linguistics #144/582)
- M. Baroni and A. Lenci. 2008. Concepts and properties in word spaces. *Italian Journal of Linguistics* 20(1). (Language and Linguistics #309/582)
- L. Onnis, T. Farmer, M. Baroni, M. Christiansen and M. Spivey. 2008. Generalizable distributional regularities aid fluent language processing: The case of semantic valence tendencies. *Italian Journal of Linguistics* 20(1). (Language and Linguistics #309/582)

- M. Baroni, E. Guevara and V. Pirrelli. 2007. NN compounds in Italian: Modelling category induction and analogical extension. *Lingue e Linguaggio*, 6.2. (Language and Linguistics #221/582)
- H. Schmid, M. Baroni, E. Zanchetta and A. Stein. 2007. The Enriched TreeTagger System. *Intelligenza Artificiale IV-2*.
- M. Baroni and S. Bernardini. 2006. A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing* 21(3). +++
- H. Trost, J. Matiassek and M. Baroni. 2005. The language component of the FASTY text prediction system. *Applied Artificial Intelligence* 19(8). (Artificial Intelligence #100/152)
- M. Baroni. 2003. Distribution-driven morpheme discovery: A computational/experimental study. *Yearbook of Morphology 2003*. (Language and Linguistics #158/582)
- M. Baroni. 2001. The representation of prefixed forms in the Italian lexicon: Evidence from the distribution of intervocalic [s] and [z] in northern Italian. *Yearbook of Morphology 1999*.
- M. Baroni. 1994. Moraic structure and vowel length in Galeatese. *Romance Linguistics and Literature Review* 7. (Language and Linguistics #158/582)
- M. Baroni. 1993. Teorie della sottospecificazione e restrizioni sulle code consonantiche in italiano. *Rivista di Grammatica Generativa* 18.

3) *Edited volumes*

- M. Baroni, S. Evert and A. Lenci (editors). 2008. Bridging the gap between semantic theory and computational simulations: Proceedings of the ESSLLI 2008 Workshop on Distributional Semantics, Hamburg: ESSLLI.
- M. Baroni, A. Lenci, and M. Sahlgren (editors). 2007. Beyond words and documents: Proceedings of the Workshop on Contextual Information in Semantic Space Models at CONTEXT 07, Roskilde: Roskilde University.
- M. Baroni and S. Bernardini (editors). 2006. Wacky! Working papers on the Web as Corpus, Bologna: Gedit.
- A. Kilgarriff and M. Baroni (editors). 2006. Proceedings of the 2nd International Workshop on the Web as Corpus (EACL 2006 SIGWAC Workshop), East Stroudsburg PA: ACL.

4) *Articles in refereed conference proceedings and refereed volume contributions*

Conference proceedings are the main tool of scientific dissemination in computational linguistics. Articles in proceedings of conferences ranked among the top 10 overall most impactful publication venues in computational linguistics (including journals) by Google Scholar are marked with 3 asterisks.

- K. Gulordava, P. Bojanowski, E. Grave, T. Linzen and M. Baroni. 2018. Colorless green recurrent networks dream hierarchically. *Proceedings of NAACL*. ***
- M. Baroni, G. Boleda and S. Pado. 2018. Show me the cup: Reference with continuous representations. *Proceedings of CICLing*. Invited article.
- A. Herbelot and M. Baroni. 2017. High-risk learning: Acquiring new word vectors from tiny data. *Proceedings of EMNLP*. ***
- A. Lazaridou, A. Peysakhovich, and M. Baroni. 2017. Multi-agent cooperation and the emergence of (natural) language. *Proceedings of ICLR*.

- G. Boleda, S. Pado, N. Pham and M. Baroni. 2017. Living a discrete life in a continuous world: Reference with distributed representations. Proceedings of IWCS.
- D. Paperno, G. Kruszewski, A. Lazaridou, N. Pham, R. Bernardi, S. Pezzelle, M. Baroni, G. Boleda and R. Fernandez. 2016. The LAMBADA dataset: Word prediction requiring a broad discourse context. Proceedings of ACL. ***
- A. Lazaridou, N. Pham and M. Baroni. 2016. The red one! On learning to refer to things based on their discriminative properties. Proceedings of ACL. ***
- A. Lazaridou, G. Chrupala, R. Fernandez and M. Baroni. 2016. Multimodal semantic learning from child-directed input. Proceedings of NAACL HLT. ***
- T. Mikolov, A. Joulin and M. Baroni. 2016. A roadmap towards machine intelligence. Proceedings of CICLing.
- A. Gupta, G. Boleda, M. Baroni and S. Pado. 2015. Distributional vectors encode referential attributes. Proceedings of EMNLP. ***
- N. Pham, G. Kruszewski, A. Lazaridou and M. Baroni. 2015. Jointly optimizing word representations for lexical and sentential tasks with the C-PHRASE model. Proceedings of ACL. ***
- N. Pham, A. Lazaridou and M. Baroni. 2015. A multitask objective to inject lexical contrast into distributional semantics. Proceedings of ACL 2015. ***
- A. Lazaridou, G. Dinu and M. Baroni. 2015. Hubness and pollution: Delving into cross-space mapping for zero-shot learning. Proceedings of ACL 2015. ***
- A. Lazaridou, N. Pham and M. Baroni. 2015. Combining language and vision with a multimodal skip-gram model. Proceedings of NAACL HLT 2015. ***
- G. Kruszewski and M. Baroni. 2015. So similar and yet incompatible: Toward automated identification of semantically compatible words. Proceedings of NAACL HLT. ***
- S. Ritter, C. Long, D. Paperno, M. Baroni, M. Botvinick and A. Goldberg. 2015. Leveraging preposition ambiguity to assess compositional distributional models of semantics. Proceedings of *SEM 2015.
- G. Kruszewski and M. Baroni. 2014. Dead parrots make bad pets: Exploring modifier effects in noun phrases. Proceedings of *SEM.
- A. Lazaridou, E. Bruni and M. Baroni. 2014. Is this a wampimuk? Cross-modal mapping between distributional semantics and the visual world. Proceedings of ACL. ***
- G. Dinu and M. Baroni. 2014. How to make words with vectors: Phrase generation in distributional semantics. Proceedings of ACL. ***
- D. Paperno, N. Pham and M. Baroni. 2014. A practical and linguistically-motivated approach to compositional distributional semantics. Proceedings of ACL. ***
- M. Baroni, G. Dinu and G. Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors Proceedings of ACL. ***
+++
- J. Li, M. Baroni and G. Dinu. 2014. Improving the lexical function composition model with pathwise optimized elastic-net regression. Proceedings of EACL. ***
- M. Marelli, S. Menini, M. Baroni, L. Bentivogli, R. Bernardi and R. Zamparelli. 2014. A SICK cure for the evaluation of compositional distributional semantic models. Proceedings of LREC. *** +++
- A. Anderson, E. Bruni, U. Bordignon, M. Poesio and M. Baroni. 2013. Of words, eyes and brains: Correlating image-based distributional semantic models with neural representations of concepts. Proceedings of EMNLP. **Best paper nomination.** ***
- E. Vecchi, R. Zamparelli and M. Baroni. 2013. Studying the recursive behaviour of adjectival modification with compositional distributional semantics. Proceedings of EMNLP. ***

- A. Lazaridou, E. Vecchi and M. Baroni. 2013. Fish transporters and miracle homes: How compositional distributional semantics can help NP parsing. Proceedings of EMNLP. ***
- A. Lazaridou, M. Marelli, R. Zamparelli and M. Baroni. 2013. Compositionally derived representations of morphologically complex words in distributional semantics. Proceedings of ACL. ***
- R. Bernardi, G. Dinu, M. Marelli and M. Baroni. 2013. A relatedness benchmark to test the role of determiners in compositional distributional semantics. Proceedings of ACL. ***
- E. Grefenstette, G. Dinu, Y.-Z. Zhang, M. Sadrzadeh and M. Baroni. 2013. Multi-step regression learning for compositional distributional semantics. Proceedings of IWCS.
- G. Boleda, M. Baroni, L. McNally and N. Pham. 2013. Intensionality was only alleged: On adjective-noun composition in distributional semantics. Proceedings of IWCS.
- M. Baroni and S. Bernardini. 2013. Corpus query tools for lexicography. In Rufus Gouws, Ulrich Heid, Wolfgang Schweickard and Herbert Wiegand (eds.), *Dictionaries: An international encyclopedia of lexicography*, Berlin: Mouton de Gruyter. Invited article.
- E. Bruni, J. Uijlings, M. Baroni and N. Sebe. 2012. Distributional semantics with eyes: Using image analysis to improve computational representations of word meaning. Proceedings of Multimedia. **Brave New Idea paper**.
- M. Baroni, R. Bernardi, N. Do and C. Shan. 2012. Entailment above the word level in distributional semantics. Proceedings of EACL. *** +++
- E. Bruni, G. Boleda, M. Baroni and N. Tran. 2012. Distributional semantics in technicolor. Proceedings of ACL. *** +++
- M. Baroni. 2011. Statistiche linguistiche. In Raffaele Simone (ed.), *Enciclopedia dell'italiano*, vol. 2. Roma: Istituto della Enciclopedia Italiana. Invited article.
- M. Baroni and R. Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. Proceedings of EMNLP. *** +++
- M. Poesio, M. Baroni, O. Lanz, A. Lenci, A. Potamianos, H. Schütze, S. Schulte im Walde and L. Surian. 2010. BabyExp: Constructing a huge multimodal resource to acquire commonsense knowledge like children do. Proceedings of LREC. ***
- M. Baroni. 2010. Corpora di italiano. In Raffaele Simone (ed.), *Enciclopedia dell'italiano*, vol. 1. Roma: Istituto della Enciclopedia Italiana. Invited article.
- V. Pirrelli, E. Guevara and M. Baroni. 2010. Computational issues in compound parsing. In Sergio Scalise and Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, Amsterdam: Benjamins
- B. Murphy, M. Baroni and M. Poesio. 2009. EEG responds to conceptual stimuli and corpus semantics. Proceedings of EMNLP. ***
- M. Baroni. 2009. Distributions in text. In Anke Lüdeling and Merja Kytö (eds.), *Corpus linguistics: An international handbook*, Volume 2, Berlin: Mouton de Gruyter. Invited article. +++
- M. Baroni and S. Evert. 2009. Statistical methods for corpus exploitation. In Anke Lüdeling and Merja Kytö (eds.), *Corpus linguistics: An international handbook*, Volume 2, Berlin: Mouton de Gruyter. Invited article.
- M. Baroni, E. Guevara and V. Pirrelli. 2009. Sulla tipologia dei composti N+N in italiano: Principi categoriali ed evidenza distribuzionale a confronto. Proceedings of SLI.
- M. Baroni, F. Chantree, A. Kilgarriff and S. Sharoff. 2008. CleanEval: A competition for cleaning Webpages. Proceedings of LREC. *** +++
- M. Baroni and S. Evert. 2007. Words and echoes: Assessing and mitigating the non-randomness problem in word frequency distribution modeling. Proceedings of ACL. ***

- M. Baroni. 2007. I sensi di ri-: Un'indagine preliminare. In Roberta Maschi, Nicoletta Penello and Piera Rizzolatti (eds.), *Miscellanea di studi linguistici offerti a Laura Vanelli*. Udine: Forum.
- A. Lüdeling, S. Evert and M. Baroni. 2007. Using Web data for linguistic purposes. In Marianne Hundt, Nadja Nesselhauf and Caroline Biewer (eds.), *Corpus linguistics and the Web*. Amsterdam: Rodopi. +++
- M. Baroni, A. Kilgarriff, J. Pomikalek and P. Rychly. 2006. WebBootCaT: a web tool for instant corpora. *Proceedings of Euralex*.
- M. Baroni, A. Kilgarriff, J. Pomikalek and P. Rychly. 2006. WebBootCaT: Instant domain-specific corpora to support human translators. *Proceedings of EAMT*.
- M. Ciaramita and M. Baroni. 2006. A figure of merit for the evaluation of Web-corpus randomness. *Proceedings of EACL*. ***
- M. Baroni and A. Kilgarriff. 2006. Large linguistically-processed Web corpora for multiple languages. *Proceedings of EACL*. *** +++
- S. Evert and M. Baroni. 2006. Testing the extrapolation quality of word frequency models. *Proceedings of Corpus Linguistics*.
- M. Ueyama and M. Baroni. 2006. Automated construction and evaluation of a Japanese web-based reference corpus. *Proceedings of Corpus Linguistics*.
- S. Bernardini and M. Baroni. 2006. Spotting translationese: A corpus-driven approach using support vector machines. *Proceedings of Corpus Linguistics*.
- E. Zanchetta and M. Baroni. 2006. Morph-it! A free corpus-based morphological resource for the Italian language. *Proceedings of Corpus Linguistics*.
- R. Scarborough, P. Keating, M. Baroni, T. Cho, S. Mattys, A. Alwan, E. Auer and L. Bernstein. 2006. Optical cues to the visual perception of lexical and phrasal stress in English. *Proceedings of Speech Prosody*.
- M. Baroni and S. Vegnaduzzo. 2004. Identifying subjective adjectives through web-based mutual information. *Proceedings of KONVENS*.
- M. Baroni and M. Ueyama. 2004. Retrieving Japanese specialized terms and corpora from the World Wide Web. *Proceedings of KONVENS*.
- M. Baroni and S. Bernardini. 2004. BootCaT: Bootstrapping corpora and terms from the web. *Proceedings of LREC*. *** +++
- M. Baroni and S. Bisi. 2004. Using cooccurrence statistics and the web to discover synonyms in a technical language. *Proceedings of LREC*. ***
- M. Baroni, S. Bernardini, F. Comastri, L. Piccioni, A. Volpi, G. Aston and M. Mazzoleni. 2004. Introducing the la Repubblica corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. *Proceedings of LREC*. *** +++
- P. Keating, M. Baroni, S. Mattys, R. Scarborough, A. Alwan, E. Auer and L. Bernstein. 2003. Optical phonetics and visual perception of lexical and phrasal stress in English. *Proceedings of ICPhS*.
- M. Baroni and S. Bernardini. 2003. A preliminary analysis of collocational differences in monolingual comparable corpora. *Proceedings of Corpus Linguistics*.
- M. Baroni, J. Matiasek and H. Trost. 2002. Wordform- and class-based prediction of the components of German nominal compounds in an AAC system. *Proceedings of COLING 2002*. ***
- M. Baroni, J. Matiasek and H. Trost. 2002. Predicting the components of German nominal compounds. *Proceedings of ECAI*.
- J. Matiasek, M. Baroni and H. Trost. 2002. FASTY: A multi-lingual approach to text prediction. *Proceedings of ICCHP*.

- M. Baroni and L. Vanelli. 2000. The relationship between vowel length and consonantal voicing in Friulian. In Lori Repetti (ed.), *Phonological theory and the dialects of Italy*. Amsterdam: John Benjamins.
- M. Baroni and L. Vanelli. 1999. Il contrasto di lunghezza vocalica in friulano. *Proceedings of SLI*.

5) *Selected manuscripts and refereed workshop articles*

- B. Lake and M. Baroni. 2018. Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks. Under review.
- A. Liska, G. Kruszewski and M. Baroni. 2018. Memorize or generalize? Searching for a compositional RNN in a haystack. Under review.
- A. Conneau, G. Kruszewski, G. Lample, L. Barrault and M. Baroni. 2018. What you can cram into a single \$&!#* vector: Probing sentence embeddings for linguistic properties. Under review.
- M. Baroni, A. Joulin, A. Jabri, G. Kruszewski, A. Lazaridou, K. Simonic and T. Mikolov. 2017. CommAI: Evaluating the first steps towards a useful general AI. *Proceedings of ICLR Workshop Track*.
- A. Lazaridou, D. Nguyen and M. Baroni. 2015. Do distributed semantic models dream of electric sheep? Visualizing word representations through image synthesis. *Proceedings of the EMNLP Workshop on Vision and Language (VL15)*.
- G. Dinu, A. Lazaridou and M. Baroni. 2015. Improving zero-shot learning by mitigating the hubness problem. *Proceedings of ICLR Workshop Track*.
- M. Marelli, L. Bentivogli, M. Baroni, R. Bernardi, S. Menini and R. Zamparelli. 2014. Semeval-2014 Task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment. *Proceedings of SemEval*. *** +++
- G. Dinu, N. Pham and M. Baroni. 2013. General estimation and evaluation of compositional distributional semantic models. *Proceedings of the ACL Workshop on Continuous Vector Space Models and their Compositionality (CVSC)*.
- E. Vecchi, M. Baroni and R. Zamparelli. 2011. (Linear) maps of the impossible: Capturing semantic anomalies in distributional space. *Proceedings of the ACL DISCO (Distributional Semantics and Compositionality) Workshop*.
- K. Gulordava and M. Baroni. 2011. A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. *Proceedings of the EMNLP Geometrical Models for Natural Language Semantics (GEMS) Workshop*.
- E. Bruni, G.B. Tran and M. Baroni. 2011. Distributional semantics from text and images. *Proceedings of the EMNLP Geometrical Models for Natural Language Semantics (GEMS) Workshop*
- M. Baroni and A. Lenci. 2011. How we BLESSEd distributional semantic evaluation. *Proceedings of the EMNLP Geometrical Models for Natural Language Semantics (GEMS) Workshop*. +++
- A. Herdagdelen and M. Baroni. 2010. The Concept Game: Better commonsense knowledge extraction by combining text mining and a game with a purpose. *Proceedings of the AAAI Fall Symposium*.
- G. Kremer and M. Baroni. 2010. Predicting cognitively salient modifiers of the constitutive parts of concepts. *Proceedings of the ACL Cognitive Modeling and Computational Linguistics Workshop*.
- M. Baroni and A. Lenci. 2009. One distributional memory, many semantic spaces. *Proceedings of the EACL Geometrical Models for Natural Language Semantics (GEMS) Workshop*.

- A. Herdagdelen and M. Baroni. 2009. BagPack: A general framework to represent semantic relations. Proceedings of the EACL Geometrical Models for Natural Language Semantics (GEMS) Workshop.
- A. Herdagdelen, K. Erk and M. Baroni. 2009. Measuring semantic relatedness with vector space models and random walks. Proceedings of TextGraphs.
- A. Ferraresi, E. Zanchetta, M. Baroni and S. Bernardini. 2008. Introducing and evaluating ukWaC, a very large Web-derived corpus of English. Proceedings of the WAC4 Workshop at LREC. **Best paper award.** +++
- M. Baroni, A. Lenci and L. Onnis. 2007. ISA meets Lara: An incremental word space model for cognitively plausible simulations of semantic learning. Proceedings of the ACL Workshop on Cognitive Aspects of Computational Language Acquisition.
- M. Baroni, J. Matiasek and H. Trost. 2002. Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. Proceedings of the ACL Workshop on Morphological and Phonological Learning. +++

Supervision

- Post-doc supervision: Germán Kruszewski (2017-present), Aurelie Herbelot (2015-2016), Gemma Boleda (2015-2016), Marco Marelli (2012-2016), Denis Paperno (2012-2016), Jiming Li (2013-2014), Georgiana Dinu (2012-2014), Yao-Zhong Zhang (2012)
- PhD supervision: Rahma Chaabouni (ongoing), German Kruszewski (2016), Angeliki Lazaridou (2016, CIMeC young scientist award), Nghia The Pham (2016), Elia Bruni (2013), Eva Maria Vecchi (co-supervision, 2013), Gerhard Kremer (2010), Amaç Herdagdelen (2010, University of Trento best dissertation award), Federico Boschetti (2009)
- External examiner of 17 international PhD candidates
- Currently managing a research team with 7 members at FAIR

Teaching

- Classes taught at the University of Trento (2007-2016): Computational linguistics (Master/BA/BSc), Computational skills for text analysis (Master/BA), Statistics practicum (PhD), Language and computation topic seminar (PhD), Humanities computing (BA), Applied linguistics (BA), General linguistics (BA)
- Curriculum coordination at the University of Trento (2009-2014): Cognitive Science Master (Language and Multimodal Interaction track), Philosophy Master and BA (Philosophy and Informatics track)
- Classes taught at the University of Bologna (2002-2006): Terminology and specialized languages (MA), Computational linguistics (BA), General linguistics (BA)
- 12 international tutorials and mini-courses offered, including at ESSLI (2006 and 2014), EACL (2012) and ACL (2013)

Last updated: February 24th, 2018