

Bibliography

- Eneko Agirre, Mona Diab, Daniel Cer, and Aitor Gonzalez-Agirre. Semeval-2012 task 6: A pilot on semantic textual similarity. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, SemEval '12, pages 385–393, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics. URL <http://dl.acm.org/citation.cfm?id=2387636.2387697>.
- Eneko Agirre, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, and Weiwei Guo. * SEM 2013 shared task: Semantic textual similarity. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity*, volume 1, pages 32–43, 2013.
- Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Weiwei Guo, Rada Mihalcea, German Rigau, and Janyce Wiebe. Semeval-2014 task 10: Multilingual semantic textual similarity. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pages 81–91, 2014.
- Eneko Agirre, Aitor Gonzalez-Agirre, Iñigo Lopez-Gazpio, Montse Maritxalar, German Rigau, and Larraitz Uria. SemEval-2016 task 2: Interpretable semantic textual similarity. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 512–524, San Diego, California, June 2016. Association for Computational Linguistics. doi: 10.18653/v1/S16-1082. URL <https://www.aclweb.org/anthology/S16-1082>.
- Hanan Aldarmaki and Mona Diab. Evaluation of unsupervised compositional representations. In *Proceedings of COLING 2018*, 2018.
- Salha Alzahrani and Naomie Salim. Fuzzy semantic-based string similarity for extrinsic plagiarism detection. *Braschler and Harman*, 1176:1–8, 2010.

- Ion Androutsopoulos and Prodromos Malakasiotis. A survey of paraphrasing and textual entailment methods. *Journal of Artificial Intelligence Research*, 38: 135–187, 2010.
- Yuki Arase and Jun’ichi Tsujii. Spade: Evaluation dataset for monolingual phrase alignment. In *Proceedings of LREC-2018*, 2018.
- Antti Arppe, Gaetanelle Gilquin, Dylan Glynn, Martin Hilpert, and Arne Zeschel. Cognitive corpus linguistics: five points of debate on current theory and methodology. *Corpora*, 5(1):1–27, 2010.
- Sören Auer, Chris Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. DBpedia: A nucleus for a web of open data. In *Proceedings of the 6th International Semantic Web Conference (ISWC)*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735. Springer, 2008. doi: [doi:10.1007/978-3-540-76298-0_52](https://doi.org/10.1007/978-3-540-76298-0_52).
- Timothy Baldwin and Su Nam Kim. Multiword Expressions. *Handbook of natural language processing*, 2:267–92, 2010.
- Daniel Bär, Torsten Zesch, and Iryna Gurevych. Text reuse detection using a composition of text similarity measures. *Proceedings of COLING 2012*, pages 167–184, 2012.
- Roy Bar-Haim, Ido Dagan, Bill Dolan, Lisa Ferro, and Danilo Giampiccolo. The second pascal recognising textual entailment challenge. In *Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*, 01 2006.
- M. Baroni, S. Bernardini, A. Ferraresi, and E. Zanchetta. The wacky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226, 2009.
- Marco Baroni. Composition in distributional semantics. *Language and Linguistics Compass*, 7(10):511–22, 2013.
- Marco Baroni and Alessandro Lenci. Distributional memory: A general framework for corpus-based semantics. *Comput. Linguist.*, 36(4):673–721, December 2010.
- Marco Baroni, Brian Murphy, Eduard Barbu, and Massimo Poesio. Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science*, 34(2):222–54, 2010.

- Alberto Barrón-Cedeño, Marta Vila, M. Antònia Martí, and Paolo Rosso. Plagiarism meets paraphrasing: Insights for the next generation in automatic plagiarism detection. *Computational Linguistics*, 39(4):917–947, 2013.
- S. Bartsch. *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Gunter Narr Verlag, 2004.
- Vuk Batanović, Miloš Cvetanović, and Boško Nikolić. Fine-grained semantic textual similarity for serbian. In *Proceedings of LREC-2018*, 2018.
- Darina Benikova and Torsten Zesch. Same same, but different: Compositionality of paraphrase granularity levels. In *Proceedings of RANLP 2017*, 2017.
- Luisa Bentivogli, Bernardo Magnini, Ido Dagan, Hoa Trang Dang, and Danilo Giampiccolo. The fifth PASCAL recognizing textual entailment challenge. In *Proceedings of the Second Text Analysis Conference, TAC 2009, Gaithersburg, Maryland, USA, November 16-17, 2009*, 2009.
- Luisa Bentivogli, Peter Clark, Ido Dagan, and Danilo Giampiccolo. The sixth PASCAL recognizing textual entailment challenge. In *Proceedings of the Third Text Analysis Conference, TAC 2010, Gaithersburg, Maryland, USA, November 15-16, 2010*, 2010.
- Luisa Bentivogli, Peter Clark, Ido Dagan, and Danilo Giampiccolo. The seventh PASCAL recognizing textual entailment challenge. In *Proceedings of the Fourth Text Analysis Conference, TAC 2011, Gaithersburg, Maryland, USA, November 14-15, 2011*, 2011.
- Manuel Bertrán, Oriol Borrega, Marta Recasens, and Bàrbara Soriano. Anco-rapipe: A tool for multilevel annotation. *Procesamiento del Lenguaje Natural*, 41, 2008. URL <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/2577/1116>.
- Rahul Bhagat. *Learning Paraphrases from Text*. PhD thesis, Los Angeles, CA, USA, 2009. AAI3368694.
- Rahul Bhagat and Eduard H. Hovy. What is a paraphrase? *Computational Linguistics*, 39(3):463–472, 2013.
- Chris Biemann and Eugenie Giesbrecht. Distributional semantics and compositionality 2011: Shared task description and results. In *Proceedings of the workshop on distributional semantics and compositionality*, pages 21–8. Association for Computational Linguistics, 2011.