

- sentence embeddings for linguistic properties. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2126–2136, Melbourne, Australia. Association for Computational Linguistics.
- Silvio Cordeiro, Aline Villavicencio, Marco Idiart, and Carlos Ramisch. 2019. Unsupervised compositionality prediction of nominal compounds. *Computational Linguistics*, 45(1):1–57.
- David Alan Cruse. 1986. *Lexical semantics*. Cambridge university press.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Kawin Ethayarajh. 2019. How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 55–65, Hong Kong, China. Association for Computational Linguistics.
- Allyson Ettinger. 2020. What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, 8:34–48.
- Meghdad Farahmand and James Henderson. 2016. Modeling the non-substitutability of multiword expressions with distributional semantics and a log-linear model. In *Proceedings of the 12th Workshop on Multiword Expressions*, pages 61–66, Berlin, Germany. Association for Computational Linguistics.
- Afsaneh Fazly, Paul Cook, and Suzanne Stevenson. 2009. Unsupervised type and token identification of idiomatic expressions. *Computational Linguistics*, 35(1):61–103.
- Christina L Gagné and Thomas L Spalding. 2009. Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language*, 60(1):20–35.
- Rachel Giora. 1999. On the priority of salient meanings: Studies of literal and figurative language. *Journal of pragmatics*, 31(7):919–929.
- Yoav Goldberg. 2019. Assessing BERT’s Syntactic Abilities. *arXiv preprint arXiv:1901.05287*.
- Kristina Gulordava, Piotr Bojanowski, Edouard Grave, Tal Linzen, and Marco Baroni. 2018. Colorless green recurrent networks dream hierarchically. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1195–1205, New Orleans, Louisiana. Association for Computational Linguistics.
- Nathan Hartmann, Erick Fonseca, Christopher Shulby, Marcos Treviso, Jéssica Silva, and Sandra Aluísio. 2017. Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology*, pages 122–131, Uberlândia, Brazil. Sociedade Brasileira de Computação.
- Hongbo Ji, Christina L Gagné, and Thomas L Spalding. 2011. Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, 65(4):406–430.
- Milton King and Paul Cook. 2017. Supervised and unsupervised approaches to measuring usage similarity. In *Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications*, pages 47–52, Valencia, Spain. Association for Computational Linguistics.
- Milton King and Paul Cook. 2018. Leveraging distributed representations and lexico-syntactic fixedness for token-level prediction of the idiomacity of english verb-noun combinations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, pages 345–350. Association for Computational Linguistics.
- Jean-Paul Laurent, Guy Denhières, Christine Passerieux, Galina Iakimova, and Marie-Christine Hardy-Baylé. 2006. On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, 1068(1):151–160.
- Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. 2016. Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, 4:521–535.
- Nelson F. Liu, Matt Gardner, Yonatan Belinkov, Matthew E. Peters, and Noah A. Smith. 2019. Linguistic knowledge and transferability of contextual representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1073–1094, Minneapolis, Minnesota. Association for Computational Linguistics.

- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. **Distributed representations of words and phrases and their compositionality**. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS’13, pages 3111–3119, USA. Curran Associates Inc.
- George A. Miller. 1995. **WordNet: a lexical database for English**. *Communications of the ACM*, 38(11):39–41.
- Navnita Nandakumar, Timothy Baldwin, and Bahar Salehi. 2019. **How Well Do Embedding Models Capture Non-compositionality? A View from Multiword Expressions**. In *Proceedings of the 3rd Workshop on Evaluating Vector Space Representations for NLP*, pages 27–34, Minneapolis, USA. Association for Computational Linguistics.
- James H. Neely, Dennis E. Keefe, and Kent L. Ross. 1989. **Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching**. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6):1003–1019.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. **Glove: Global vectors for word representation**. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.
- Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. **Deep contextualized word representations**. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.
- Grusha Prasad, Marten van Schijndel, and Tal Linzen. 2019. **Using priming to uncover the organization of syntactic representations in neural language models**. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 66–76, Hong Kong, China. Association for Computational Linguistics.
- Siva Reddy, Diana McCarthy, and Suresh Manandhar. 2011. **An empirical study on compositionality in compound nouns**. In *Fifth International Joint Conference on Natural Language Processing, IJCNLP 2011, Chiang Mai, Thailand, November 8-13, 2011*, pages 210–218. The Association for Computer Linguistics.
- Nils Reimers and Iryna Gurevych. 2019a. **Alternative Weighting Schemes for ELMo Embeddings**. *CoRR*, abs/1904.02954.
- Nils Reimers and Iryna Gurevych. 2019b. **Sentence-BERT: Sentence embeddings using Siamese BERT-networks**. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.
- Joost Rommers, Ton Dijkstra, and Marcel Bastiaansen. 2013. **Context-dependent semantic processing in the human brain: Evidence from idiom comprehension**. *Journal of Cognitive Neuroscience*, 25(5):762–776.
- Andreas Rücklé, Steffen Eger, Maxime Peyrard, and Iryna Gurevych. 2018. **Concatenated power mean word embeddings as universal cross-lingual sentence representations**. *arXiv preprint arXiv:1803.01400*.
- Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. **Multiword expressions: A pain in the neck for NLP**. In *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2002)*, pages 1–15, Mexico City, Mexico. Springer, Berlin, Heidelberg.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. **DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter**. In *Proceedings of the 5th Workshop on Energy Efficient Machine Learning and Cognitive Computing at NeurIPS 2019*, Vancouver, Canada.
- Timo Schick and Hinrich Schütze. 2019. **Attentive mimicking: Better word embeddings by attending to informative contexts**. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 489–494, Minneapolis, Minnesota. Association for Computational Linguistics.
- Timo Schick and Hinrich Schütze. 2020a. **BERTRAM: Improved word embeddings have big impact on contextualized model performance**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3996–4007, Online. Association for Computational Linguistics.
- Timo Schick and Hinrich Schütze. 2020b. **Rare Words: A Major Problem for Contextualized Embeddings and How to Fix it by Attentive Mimicking**. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 8766–8774.
- Marten van Schijndel and Tal Linzen. 2018. **A neural model of adaptation in reading**. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4704–4710, Brussels, Belgium. Association for Computational Linguistics.

Tal Schuster, Ori Ram, Regina Barzilay, and Amir Globerson. 2019. [Cross-lingual alignment of contextual word embeddings, with applications to zero-shot dependency parsing](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1599–1613, Minneapolis, Minnesota. Association for Computational Linguistics.

Weijia Shi, Muhan Chen, Pei Zhou, and Kai-Wei Chang. 2019. [Retrofitting contextualized word embeddings with paraphrases](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1198–1203, Hong Kong, China. Association for Computational Linguistics.

Vered Shwartz and Ido Dagan. 2019. [Still a pain in the neck: Evaluating text representations on lexical composition](#). *Transactions of the Association for Computational Linguistics*, 7:403–419.

Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R Bowman, Dipanjan Das, and Ellie Pavlick. 2019. [What do you learn from context? Probing for sentence structure in contextualized word representations](#). In *Proceedings of the Seventh International Conference on Learning Representations (ICLR 2019)*, New Orleans, Louisiana.

Jorge A. Wagner Filho, Rodrigo Wilkens, Marco Idiart, and Aline Villavicencio. 2018. [The brWaC corpus: A new open resource for Brazilian Portuguese](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Gregor Wiedemann, Steffen Remus, Avi Chawla, and Chris Biemann. 2019. [Does BERT Make Any Sense? Interpretable Word Sense Disambiguation with Contextualized Embeddings](#). In *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019): Long Papers*, pages 161–170, Erlangen, Germany. German Society for Computational Linguistics & Language Technology.

Appendices

A Naturalistic examples in English

Table 6 includes naturalistic examples in English. We include the compositionality scores provided by the annotators and the BERT and GloVe results at NC level.

B Naturalistic examples in Portuguese

Table 7 includes naturalistic examples in Portuguese. We include the compositionality scores provided by the annotators and the BERT and GloVe results at NC level.