

- Gagné, Christina L. 2001. Relation and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1):236. <https://doi.org/10.1037/0278-7393.27.1.236>, PubMed: 11204100
- Gauthier, Jon and Roger Levy. 2019. Linking artificial and human neural representations of language. 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 529–539. <https://doi.org/10.18653/v1/D19-1050>
- Girju, Roxana, Ana-Maria Giuglea, Marian Olteanu, Ovidiu Fortu, Orest Bolohan, and Dan Moldovan. 2004. Support vector machines applied to the classification of semantic relations in nominalized noun phrases. In *Proceedings of the Computational Lexical Semantics Workshop at HLT-NAACL 2004*, pages 68–75. <https://doi.org/10.3115/1596431.1596441>
- Kriegeskorte, Nikolaus, Marieke Mur, and Peter Bandettini. 2008. Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:4. <https://doi.org/10.3389/neuro.06.004.2008>, PubMed: 19104670
- Lample, Guillaume and Alexis Conneau. 2019. Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291*.
- Lees, Robert B. 1960. The grammar of English nominalizations. *International Journal of American Linguistics*, 26(3):205.
- Levi, Judith N. 1978. *The Syntax and Semantics of Complex Nominals*. Academic Press, New York.
- Li, Siyan, Riley Carlson, and Christopher Potts. 2022. Systematicity in GPT-3’s interpretation of novel English noun compounds. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 717–728. <https://doi.org/10.18653/v1/2022.findings-emnlp.50>
- Linzen, T. A. L. 2019. What can linguistics and deep learning contribute to each other? Response to Pater. *Language*, 95(1):e99–e108. <https://doi.org/10.1353/lan.2019.0015>
- Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692 [cs]*.
- Lynott, Dermot and Louise Connell. 2010. Embodied conceptual combination. *Frontiers in Psychology*, 1:212. <https://doi.org/10.3389/fpsyg.2010.00212>, PubMed: 21833267
- Maguire, Phil, Barry Devereux, Fintan J. Costello, and Arthur Cater. 2007. A re-analysis of the CARIN theory of conceptual combination. *Journal of Experimental Psychology, Learning Memory and Cognition*, 33:811–821. <https://doi.org/10.1037/0278-7393.33.4.811>, PubMed: 17576155
- McCoy, R. Thomas, Junghyun Min, and Tal Linzen. 2020. BERTs of a feather do not generalize together: Large variability in generalization across models with similar test set performance. In *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, pages 217–227. <https://doi.org/10.18653/v1/2020.blackboxnlp-1.21>
- Merity, Stephen, Caiming Xiong, James Bradbury, and Richard Socher. 2016. Wikitext-103. Technical report, Salesforce.
- Mitchell, Jeff and Mirella Lapata. 2010. Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388–1429. <https://doi.org/10.1111/j.1551-6709.2010.01106.x>, PubMed: 21564253
- Mitchell, Tom M., Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. 2008. Predicting human brain activity associated with the meanings of Nouns. *Science*, 320(5880):1191–1195. <https://doi.org/10.1126/science.1152876>, PubMed: 18511683
- Murphy, Gregory L. 1988. Comprehending complex concepts. *Cognitive Science*, 12(4):529–562. [https://doi.org/10.1016/0364-0213\(88\)90012-2](https://doi.org/10.1016/0364-0213(88)90012-2)
- Murphy, Gregory L. 2002. *The Big Book of Concepts*. MIT Press, Boston, MA. <https://doi.org/10.7551/mitpress/1602.001.0001>
- Murty, Shikhar, Pratyusha Sharma, Jacob Andreas, and Christopher D. Manning. 2022. Characterizing intrinsic compositionality in transformers with tree projections. *arXiv preprint arXiv:2211.01288*.
- Nakov, Preslav. 2019. Paraphrasing verbs for noun compound interpretation. *arXiv preprint arXiv:1911.08762*.

- Nili, Hamed, Cai Wingfield, Alexander Walther, Li Su, William Marslen-Wilson, and Nikolaus Kriegeskorte. 2014. A toolbox for representational similarity analysis. *PLoS Computational Biology*, 10(4):e1003553. <https://doi.org/10.1371/journal.pcbi.1003553>, PubMed: 24743308
- Ontanon, Santiago, Joshua Ainslie, Zachary Fisher, and Vaclav Cvicek. 2022. Making transformers solve compositional tasks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3591–3607. <https://doi.org/10.18653/v1/2022.acl-long.251>
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9.
- Reddy, Siva, Diana McCarthy, and Suresh Manandhar. 2011. An empirical study on compositionality in compound nouns. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 210–218.
- Rogers, Anna, Olga Kovaleva, and Anna Rumshisky. 2021. A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics*, 8:842–866. https://doi.org/10.1162/tacl_a_00349
- Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- Sellam, Thibault, Steve Yadlowsky, Jason Wei, Naomi Saphra, Alexander D’Amour, Tal Linzen, Jasmijn Bastings, Julia Turc, Jacob Eisenstein, and Dipanjan Das. 2021. The multiBERTs: BERT reproductions for robustness analysis. *arXiv preprint arXiv:2106.16163*.
- Shwartz, Vered and Ido Dagan. 2018. Paraphrase to explicate: Revealing implicit noun-compound relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1200–1211. <https://doi.org/10.18653/v1/P18-1111>
- Shwartz, Vered 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. https://doi.org/10.1162/tacl_a_00277
- Tenney, Ian, Dipanjan Das, and Ellie Pavlick. 2019. BERT rediscovers the classical NLP pipeline. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4593–4601. <https://doi.org/10.18653/v1/P19-1452>
- Tratz, Stephen and Eduard Hovy. 2010. A taxonomy, dataset, and classifier for automatic noun compound interpretation. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 678–687.
- van Jaarsveld, Henk J. and Gilbert E. Rattink. 1988. Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of Psycholinguistic Research*, 17(6):447–473. <https://doi.org/10.1007/BF01067911>
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances In Neural Information Processing Systems*.
- Westerlund, Masha and Liina Pylkkänen. 2017. How does the left anterior temporal lobe contribute to conceptual combination? Interdisciplinary perspectives. In *Compositionality and Concepts in Linguistics and Psychology*. Springer, Cham, pages 269–290. https://doi.org/10.1007/978-3-319-45977-6_11
- Wicherts, Jelte M., Coosje L. M. Veldkamp, Hilde E. M. Augusteijn, Marjan Bakker, Robbie Van Aert, and Marcel A. L. M. Van Assen. 2016. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, page 1832. <https://doi.org/10.3389/fpsyg.2016.01832>, PubMed: 27933012
- Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierrick Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>
- Xu, Haoyan, Brian Murphy, and Alona Fyshe. 2016. BrainBench: A brain-image

- test suite for distributional semantic models. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2017–2021. <https://doi.org/10.18653/v1/D16-1213>
- Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R. Salakhutdinov, and Quoc V. Le. 2019. XLNet: Generalized autoregressive pretraining for language understanding. *Advances In Neural Information Processing Systems*, 32.
- Yu, Lang and Allyson Ettinger. 2020. Assessing phrasal representation and composition in transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4896–4907. <https://doi.org/10.18653/v1/2020.emnlp-main.397>
- Yu, Lang and Allyson Ettinger. 2021. On the interplay between fine-tuning and composition in transformers. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2279–2293. <https://doi.org/10.18653/v1/2021.findings-acl.201>
- Ó Séaghdha, Diarmuid. and Ann Copestake. 2007. Co-occurrence contexts for noun compound interpretation. In *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions*, pages 57–64. <https://doi.org/10.3115/1613704.1613712>