

Zhiguo Wang, Haitao Mi, and Abraham Ittycheriah. Sentence similarity learning by lexical decomposition and composition. *CoRR*, abs/1602.07019, 2016. URL <http://arxiv.org/abs/1602.07019>.

David Wible and Nai-Lung Tsao. StringNet As a Computational Resource for Discovering and Investigating Linguistic Constructions. In *Proceedings of the NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics*, EUCCCL '10, pages 25–31, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.

Adina Williams, Nikita Nangia, and Samuel Bowman. A broad-coverage challenge corpus for sentence understanding through inference. 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 1112–1122, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1101. URL <https://www.aclweb.org/anthology/N18-1101>.

Ludwig Wittgenstein. *Philosophical Investigations*. (Translated by Anscombe, G.E.M.). Basil Blackwell, 1953.

Alison Wray and Mick Perkins. The functions of formulaic language: an integrated model. *Language and Communication*, 20(1):1–28, 2000.

Ronald Yager. Default knowledge and measures of specificity. 61:1–44, 04 1992.

Seid Muhie Yimam and Chris Biemann. Par4Sim – adaptive paraphrasing for text simplification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 331–342, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/C18-1028>.

Seid Muhie Yimam and Iryna Gurevych. Webanno: A flexible, web-based and visually supported system for distributed annotations. In *In Proceedings of ACL-2013 System Demonstrations*, pages 1–6, 2013.

Ken-ichi Yokote, Shohei Tanaka, and Mitsuru Ishizuka. Effects of Using Simple Semantic Similarity on Textual Entailment Recognition. In *TAC*, 2011.

Willem Zuidema. What are the productive units of natural language grammar?: a dop approach to the automatic identification of constructions. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, pages 29–36. Association for Computational Linguistics, 2006.



# Appendix A

## Annotation Guidelines for ETPC

### A.1 Presentation

This document sets out the guidelines for the paraphrase typology annotation task, using the Extended Paraphrase Typology. The task consists of annotating candidate paraphrase pairs (including positive and negative examples of paraphrasing) with a textual paraphrase label, the paraphrase types they contain, and negation. These guidelines have been used to annotate the Microsoft Research Paraphrase Corpus (MRPC), thus giving raise to the Extended Typology Paraphrase Corpus (ETPC). For the purpose of the annotation, we have developed a web based annotation tool, the WARP-Text interface.

This document is divided in five blocks: general considerations about the task and theoretical definitions (Section A.2); tagset definition (section A.3); guidelines for annotating non-paraphrases (section A.4); annotating negation (section A.5).

Marks and symbols used in this document:

- Fragments in the examples that should be annotated are underlined.  
When no fragment is underlined, it means that it is the whole example that should be tagged.
- The so-called “key elements” are in **bold**.

#### A.1.1 Credits

This document has been adapted and extended from the paraphrase typology annotation guidelines of Vila and Marti (2012).