

# Using Wordnet Senses to Investigate Zipf’s meaning-frequency Law

FRANCIS BOND<sup>1</sup>, ARKADIUSZ JANZ<sup>2</sup>, MAREK MAZIARZ<sup>2</sup>, AND EWA RUDNICKA<sup>2</sup>

<sup>1</sup>*Palacký University, Czechia, bond@ieee.org*

<sup>2</sup>*Wrocław University of Science and Technology,  
Poland, {arkadius.janz|marek.maziarz|ewa.rudnicka}@pwr.edu.pl*

**Keywords:** Zipf’s law, wordnets, sense-annotated corpora

George K. Zipf [10] observed that more frequent words have more senses. We take advantage of new sense-annotated corpora to test this law using corpora and wordnets of English, Spanish, Portuguese, French, Polish, Japanese, Indonesian and Chinese [1, 2, 3, 4, 5, 6, 7, 8, 9]. We show that the law describes languages accurately for these languages if we take - as Zipf did - mean values of meaning count and averaged ranks. The law only works over aggregates, it does not predict the number of senses for a single lemma. We also provide evidence that slope coefficients of Zipfian log-log linear model may vary from language to language. Finally, we show that the law holds for both single and multi-word expressions.

## Bibliography

- [1] Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226.
- [2] Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In *ACL 2013*, pp 1352–1362.
- [3] Francis Bond, Hitoshi Isahara, Kyoko Kanzaki, and Kiyotaka Uchimoto. 2008. LREC-2008.
- [4] Francis Bond, Lian Tze Lim, Enya Kong Tang, and Riza Hammam. 2014. The combined Wordnet Bahasa. *NUSA: Linguistic studies of languages in and around Indonesia*, 57:83–100.
- [5] Francis Bond and Liling Tan. 2012. Building and annotating the linguistically diverse NTU-MC (NTU-multilingual corpus). *Glottometrics*, 22(4):161–174.
- [6] Christiane Fellbaum, editor. 1998. *WordNet: an electronic lexical database*. MIT Press.
- [7] Aitor Gonzalez-Agirre, Egoitz Laparra, and German Rigau. 2012. Multilingual Central Repository version 3.0. In *LREV 2012*, pp 2525–2529.
- [8] Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kędzia. 2016. plWordNet 3.0 — a Comprehensive Lexical–Semantic Resource. In *COLING 2016*, pp 2259–2268.
- [9] Shan Wang and Francis Bond. 2013. Building the Chinese Open Wordnet (COW): Starting from core synsets. In *11th Workshop on Asian Language Resources*, pp 10–18.
- [10] George Kingsley Zipf. 1945. The meaning-frequency relationship of words. *The Journal of General Psychology*, 33(2):251–256.