

# More is better: Language statistics reveal a bias towards addition

Bodo Winter<sup>1</sup>, Martin H. Fischer<sup>2</sup>, Christoph Scheepers<sup>3</sup> & Andriy Myachykov<sup>3</sup>

<sup>1</sup>University of Birmingham, b.winter@bham.ac.uk <sup>2</sup>University of Potsdam, <sup>3</sup>University of Glasgow,

<sup>4</sup>Northumbria University; Institute for Cognitive Neuroscience, Higher School of Economics

**Keywords:** numerical cognition, corpus analysis, word frequency, distributional semantics

One of the core tenets of cognitive linguistics is that language is not a separate cognitive system, but connected to domain-general cognitive processes (e.g., Bybee & McClelland 2005). Language statistics in particular have been shown to reflect embodied relations, e.g., Louwerse (2008) shows that spatial iconicity is reflected in ordering preferences for binomial expressions, with pairs such as attic and basement being more frequent than the reverse, basement and attic. Similarly, word frequency statistics show that the English language is biased towards vision, in line with the Western cultural bias towards this sensory modality (Winter, Perlman & Majid 2018).

Here, we investigate how language statistics reveal a recently uncovered bias: people's tendency to neglect subtractive solutions to problems (Adams et al., 2021). For example, when tasked to "improve" a sandwich recipe, people will tend to add rather than subtract ingredients, and when tasked to review papers, reviewers will tend to make additive rather than subtractive suggestions.

This paper demonstrates that the English language shows this addition bias across numerous levels of linguistic analysis. We first use the Corpus of Contemporary American English to show that words diagnostic of addition or the concept of 'more' (add, addition, plus, more, ...) are more frequent than their subtractive counterparts (subtract, subtraction, minus, less, ...). For our statistical analysis, we fitted a Bayesian negative binomial regression model (Winter & Bürkner 2021) with the fixed effect type (add vs. subtract) and random effects for word, register, and text file, including random slopes. In a second analysis, we use the same statistical modelling approach to show that English binomial expressions reveal addition bias in their ordering preferences, e.g., add and subtract is more frequent than subtract and add etc. Finally, we show that addition-related words are also more positive in terms of their connotation, as established via corpus-based contextual valence (Snefjella & Kuperman 2016).

Finally, we use distributional semantic word vectors (word2vec, Mikolov et al. 2013) to show that synonyms of to change and to improve (as determined via thesaurus.com) are closer in semantic space to addition-related concepts. This analysis shows that the behavioral profile of words like to transform, to remodel, or to alter is similar to words such as to add and to increase. We then use a state-of-the-art language model, GPT-3, to demonstrate that in contexts such as I suggest we change this by adding/removing, adding has a higher contextual probability than removing for all verbs of change we consider (linear mixed model fitted on log probabilities with random effect for word). Altogether, our results point to an addition bias being deeply rooted in language, including word frequency, word order preferences, emotional connotation, and the usage-based semantics of verbs of change in the lexicon. Our analyses also make a methodological contribution, showing how readily available rating datasets and off-the-shelf word vectors can aid in cognitive linguistic research, specifically with respect to research that looks at how language use relates to cognition more widely.

## References

- Bybee, Joan & James L. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*. De Gruyter Mouton 22(2–4). 381–410. <https://doi.org/10.1515/tlir.2005.22.2-4.381>.
- Louwerse, Max. 2008. Embodied relations are encoded in language. *Psychonomic Bulletin & Review* 15(4). 838–844. <https://doi.org/10.3758/PBR.15.4.838>.
- Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Snefjella, Bryor & Victor Kuperman. 2016. It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition* 156. 135–146.
- Winter, Bodo & Paul-Christian Bürkner. 2021. Poisson regression for linguists: A tutorial introduction to modeling count data with brms. *Language and Linguistics Compass* 15(11). e12439. <https://doi.org/10.1111/lnc3.12439>.
- Winter, Bodo, Marcus Perlman & Asifa Majid. 2018. Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition* (179). 213–220.