COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13254

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13254

More is Better: English Language Statistics are Biased Toward Addition

Bodo Winter,[a] Martin H. Fischer,[b] Christoph Scheepers,[c] Andriy Myachykov[d,e]

[a] *Department of English Language & Linguistics, University of Birmingham*
[b] *Division of Cognitive Sciences, University of Potsdam*
[c] *School of Neuroscience and Psychology, University of Glasgow*
[d] *Department of Psychology, Northumbria University*
[e] *Institute for Cognitive Neuroscience, Higher School of Economics*

Received 15 April 2022; received in revised form 10 November 2022; accepted 30 December 2022

Abstract

We have evolved to become who we are, at least in part, due to our general drive to create new things and ideas. When seeking to improve our creations, ideas, or situations, we systematically overlook opportunities to perform subtractive changes. For example, when tasked with giving feedback on an academic paper, reviewers will tend to suggest additional explanations and analyses rather than delete existing ones. Here, we show that this addition bias is systematically reflected in English language statistics along several distinct dimensions. First, we show that words associated with an increase in quantity or number (e.g., *add, addition, more, most*) are more frequent than words associated with a decrease in quantity or number (e.g., *subtract, subtraction, less, least*). Second, we show that in binomial expressions, addition-related words are mentioned first, that is, *add and subtract* rather than *subtract and add*. Third, we show that the distributional semantics of verbs of change, such as *to improve* and *to transform*, overlap more with the distributional semantics of *add/increase* than *subtract/decrease*, which suggests that change verbs are implicitly biased toward addition. Fourth, addition-related words have more positive connotations than subtraction-related words. Fifth, we demonstrate that state-of-the-art large language models, such as the Generative Pre-trained Transformer

Correspondence should be sent to Bodo Winter, Department of English Language & Linguistics, University of Birmingham, Frankland Building, B15 2TT Edgbaston, UK. E-mail: bodo@bodowinter.com

(GPT-3), are also biased toward addition. We discuss the implications of our results for research on cognitive biases and decision-making.

Keywords: Addition; Subtraction; Subtraction neglect; Latent semantic analysis; Word frequency; Heuristics and biases

> The word “add” is a positive word. Adding something to something else usually makes it better. For example, if you add sugar to your coffee, it will probably taste better. If you add a new friend to your life, you will probably be happier.
>
> - GPT3

1. Introduction

Existing research demonstrates that people more often talk about what is important or noteworthy to them. For example, people talk more about animals that are culturally significant or visually salient (Ladle, Jepson, Correia, & Malhado, 2019), about diseases when there have been recent outbreaks (Michel et al., 2011), about ice and snow when living in colder climates (Regier, Carstensen, & Kemp, 2016), and about visual experiences in vision-dominant cultures (Winter, Perlman, & Majid, 2018). Another example of a tendency to prioritize culturally salient conceptual features is the tendency of English speakers to verbalize positively valenced words more frequently than negatively valenced ones (Augustine, Mehl, & Larsen, 2011; Boucher & Osgood, 1969; Kloumann, Danforth, Harris, Bliss, & Dodds, 2012; Warriner & Kuperman, 2015; Zajonc, 1968, p. 2), which is thought to reflect a general prosocial bias for communicative acts to be benevolent in nature (Warriner & Kuperman, 2015). Together, these findings suggest that word frequency statistics reflect what is important to speakers.

Language reflects cultural preoccupations not only via the overall frequencies of words, but also via how words pattern together with other words. Following Firth’s (1957, p. 179) credo that “you shall know a word by the company it keeps,” distributional semantics quantifies the similarity between words by looking at the similarity of the linguistic contexts in which they occur (Günther, Rinaldi, & Marelli, 2019; Hilpert & Saavedra, 2020; Landauer & Dumais, 1997; Lenci, 2008; Lund & Burgess, 1996). Just like word frequencies, distributional semantics can reveal biases that exist within a culture (Caliskan, Bryson, & Narayanan, 2017), or across cultures: In their analysis of 25 different languages, Lewis and Lupyan (2020) showed that people’s implicit gender biases, measured in a behavioral task, are predicted by language statistics; such as, how much occupation terms like *nurse* and *philosopher* overlap with gender-specific language in their distributional semantics (*he/she*, *male/female*, *boy/girl*, *man/woman*, etc.). Similarly, the fact that colors have culture-specific associations with gender in purely perceptual tasks (Jonauskaite et al., 2019) is reflected in whether color terms have semantic overlap with male- or female-biased words (Jonauskaite, Sutton, Cristianini, & Mohr, 2021). Even moral beliefs are reflected in the semantics extracted from corpora (Jentzsch, Schramowski, Rothkopf, & Kersting, 2019). Together, these studies indicate that the pattern of word usage in context can reveal biases in human cognition.

The fact that language reflects our conceptualization of the world has also been explored in discussions surrounding embodied approaches to cognition. For example, Zwaan and Yaxley (2003) showed that responses to word pairs presented on the screen were faster when in congruent vertical position (the word *attic* above *basement*) than in an incongruent one (the word *basement* above *attic*). This effect was first attributed to purely "embodied" effects, that is, visual-spatial representations becoming active during the processing of the noun pairs. Subsequently, Louwerse (2008) showed that the processing advantage of spatially congruent displays could in part be attributed to the order in which the corresponding words feature in binomial expressions, for example, *attic and basement* tends to be more frequent than the reverse order *basement and attic*. Findings such as these led Louwerse (2011) to propose a "symbol interdependency" theory according to which findings previously attributed to purely embodied effects (e.g., people performing simulations of spatial arrangements) may also have a linguistic component, given that language statistics reflect "embodied" relations (Louwerse, 2008). For example, the fact that taste and smell are highly associated sensory modalities in perception is reflected in the fact that taste and smell words are associated with language (Louwerse, 2018; Louwerse & Connell, 2011; Winter, 2016, 2019). New research suggests that even perceptual color spaces can be partially reconstructed from large language models (Abdou et al., 2021).

In this paper, we use language statistics to investigate whether there is a cultural preference for addition and "more" over subtraction and "less." Our analysis is motivated by Adams, Converse, Hales, and Klotz (2021), who provided behavioral evidence that people systematically prefer adding elements to existing objects, ideas, or situations for "improvement," and that they generally fail to consider opportunities to subtract elements. For example, when asked to improve recipes, participants are more likely to add ingredients rather than remove them. Or when asked to make suggestions for improvement within their university, people are more likely to favor adding new systems and policies than removing existing ones. Even rather abstract and decontextualized tasks reveal a bias toward additive solutions. To demonstrate this, Adams et al. (2021) presented participants with asymmetrical visual displays and asked them to produce symmetry either by adding or removing elements. Overall, very few participants even considered that they could also remove elements from the visual display to make things symmetrical. Instead, most participants produced symmetry via adding elements. These results suggest that there is a general bias toward additive solutions to problems and tasks, or on the flipside, a general neglect of subtractive solutions, which may have deep-rooted evolutionary as well as cultural origins (see Klotz, 2021).

If, as argued above, language statistics reflect existing biases, we should expect that language also reflects the "subtraction neglect" bias observed in the behavioral data. Our general assumption is that just as there is a behavioral bias toward addition, as evidenced by the studies conducted by Adams et al. (2021), there is a comparable linguistic bias. If such a linguistic bias existed, this would show that culture writ large—seen through the lens of language statistics—reflects a preoccupation with addition and "more" at the expense of subtraction and "less," in line with the ideas presented by Klotz (2021). We considered multiple different dimensions of this bias, including (1) word frequency, (2) binomial expressions, (3) semantic prosody, (4) distributional semantics using word2vec embeddings, and (5) sequential word probabilities of the state-of-the-art deep learning model GPT-3. Each of these approaches will be explained in turn.

2. This study: Multiple dimensions of addition bias

The first part focuses on a small set of words that are highly diagnostic of addition and subtraction, or more generally, the ideas of "more" as opposed to "less." These words were also chosen to limit polysemy, specifically, contexts that are not used with some invocation of quantity concepts. This has the advantage that the items included in the analyses are clearly and unambiguously related to addition or subtraction. A disadvantage of focusing on such a small set is that it is not clear to which extent the reported results generalize beyond the selected items. Therefore, a second analysis uses distributional semantics to look at a larger set of words. Distributional semantics represents words as vectors in a semantic space that is inferred from textual co-occurrences (for introduction, see Günther et al., 2019; Heylen, Wielfaert, Speelman, & Geeraerts, 2015). Such word vectors have successfully been used to predict lexical ambiguity structure (Beekhuizen, Armstrong, & Stevenson, 2021), word association judgments, synonymy, analogy, and the selection preferences of words (Baroni, Dinu, & Kruszewski, 2014; Pereira, Gershman, Ritter, & Botvinick, 2016). The utility of word space models has also been demonstrated in lexicographical analyses of polysemy (Heylen et al., 2015) and in research on grammaticalization (Hilpert & Saavedra, 2020). Here, we investigate whether proximity to addition-related words in distributional-semantic space predicts a word's frequency and emotional valence, that is, whether a word is positive or negative.

We start out with the simplest way to investigate a linguistic bias toward addition, which is to look at the occurrence frequencies of words that are strongly associated with addition and compare it to words strongly associated with subtraction. In fact, Zajonc (1968, pp. 4–5) already showed that the English word *add* is more frequent than *subtract*. In our first analysis, we build on this insight and generalize this anecdotal observation for a single item to a small set of words that are diagnostic of addition versus subtraction. We also replicate this result with more modern corpora that span a more diverse range of registers, and we use modern statistical approaches that are able to take the hierarchical nature of corpora into account (Winter & Grice, 2021).

Second, we looked at English binomial expressions, such as *salt and pepper* and *ladies and gentlemen*. There are many different factors that determine the ordering of terms in such conjoined binomial expressions; among them are semantic factors (Benor & Levy, 2006), according to which members of the pair are "ordered in accordance with a hierarchy of values inherent in the structure of a given society" (Malkiel, 1959, p. 145). The fact that binomial expressions can reveal real-world biases is also demonstrated by Louwerse's (2008) study mentioned above showing that the vertically higher member of a pair (*attic* as opposed to *basement*) tends to be mentioned first, for example, *attic and basement*. Thus, in line with a linguistic bias toward addition, we expect that addition-related terms are mentioned first in these types of expressions, for example, *add and subtract* as opposed to *subtract and add*.

Third, we examined emotional valence, an important component of many lexical representations (Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Houwer & Randell, 2004; Kuperman, Estes, Brysbaert, & Warriner, 2014; Vigliocco, Meteyard, Andrews, & Kousta, 2009). Specifically, we were interested in establishing whether "more" is also generally "better" across words. That is, we seek to establish whether addition-related words are more positive overall than subtraction-related words in their usage. This may be expected because people value "more" for cultural reasons (e.g., more income), but also because of the mere exposure effect, following from the frequency result: things that are more frequent are generally more liked (Zajonc, 1968). Another reason to expect addition-related words to be more positively valenced is that people also prefer objects of relatively larger as opposed to a smaller size (Meier, Robinson, & Caven, 2008).

Fourth, we investigated whether words of change that do not explicitly signal addition/subtraction, such as *to improve* and *to transform*, may be biased toward addition. This is a key transition point in our manuscript where we move beyond the small set of highly diagnostic words to demonstrate a more general bias that has widespread manifestations in the English language. Specifically, we investigated the distributional semantics of verbs of change (*to change, to improve, to transform*, etc.). We decided to focus on these verbs because they allow us to specifically look at whether there is a linguistic bias with respect to making suggestions for changes, which closely corresponds to the behavioral studies of Adams et al. (2021), where participants were asked to "change" or "improve" things. For this analysis, we use distributional semantics. As is done in distributional-semantic research on such topics as gender biases (e.g., Jonauskaite et al., 2021; Lewis & Lupyan, 2020), we computed each word vector's similarity to the vectors of target words, in this case, words that denote addition or subtraction. If verbs of change have meanings biased toward addition rather than subtraction, this would suggest that when people use language to describe change, or instruct people to undertake changes, this language will lead people to consider additive solutions more than subtractive solutions (cf. Fischer et al., 2021). After showing that this analysis works for a small set of synonyms of *to change* and *to improve*, we extend it to a wider set of 8000 words. The latter shows that, as words are closer to *add/addition* in semantic space, they are also much more frequent, and slightly more positively valenced.

We also include a final set of analyses where we use OpenAI's GPT-3 (Generative Pretrained Transformer 3) large language model (Brown et al., 2020). This deep learning model was designed to produce naturalistic text. The model was trained on a large collection of texts to tune a total of 175 billion parameters. As part of its sequential text generation process, GPT-3 assigns probabilities to specific words, depending on their contextual probability. We use this to assess whether GPT-3 is more likely to expect positive words such as *good* and *bad* in input sentences that contain addition- as opposed to subtraction-related words. Furthermore, we examine probe contexts such as *I suggest we change this by adding/removing* to assess whether for different verbs of change, the word *adding* is assigned a higher probability than the word *removing*, thereby replicating our word2vec analysis with cloze probabilities.

3. Method

3.1. Statistical analyses

3.1.1. Statistical tools

All statistical analyses were conducted with R version 4.0.2 (R Core Team, 2019). We used the tidyverse package version 1.3.1 (Wickham et al., 2019) and the patchwork package version 1.1.1 (Pedersen, 2020) for data processing and visualization. We used the brms package version 2.15.0 (Bürkner, 2017) for Bayesian modeling. The tidybayes package version 3.0.1 was used for plotting models (Kay, 2021). The effsize package version 0.8.1 (Torchiano, 2019) was used to calculate Cohen's *d*. The lsa package version 0.73.2 (Wild, 2020) was used to calculate cosine similarities between word vectors.

We used Bayesian analyses for several reasons. First, for complex models, especially those involving both complex random effects structures and non-normal distributions, Bayesian models estimated via Markov Chain Monte Carlo are much more likely to converge than corresponding frequentist models fitted with the widely used lme4 package (Bates, Maechler, Bolker, & Walker, 2015). A second reason is that we deal with several non-normal models in this analysis which require approaches such as negative binomial regression (Winter & Bürkner, 2021). The models we used to characterize the underlying data-generating processes are not implemented in lme4, but in brms (see Bürkner, 2017, 2018, for a discussion of supported model types). Third and finally, Bayesian models allow us to incorporate "mild skepticism" (McElreath, 2020) into our models via the incorporation of "regularizing" or "weakly informative" priors (Lemoine, 2019). These priors bias fixed effects coefficients slightly toward zero. This is especially important given that several of our analyses deal with relatively small datasets, for which estimates are known to be inherently more variable. The specific prior choices are explained for the respective analyses below, and more information about prior choices can also be found in the OSF repository (https://osf.io/detuy/), which give more detailed explanations for why we chose particular priors for each specific analysis.

3.1.2. Stimuli

We analyze the target words in Table 1, which is a hand-selected group of words directly connected to addition or subtraction. While other words can also be used in quantitative contexts, for example, *to rise* and *to fall*, or *to grow* and *to shrink*, many of these words also have alternative uses, for example, in reference to vertical position or size rather than quantity. The words in Table 1 all have a quantity-related sense that corresponds to adding or subtracting as their primary meaning. While our first analysis relies heavily on this set of hand-selected words, our analysis that uses distributional semantics extends beyond the words just considered in Table 1, thereby affording more generalizability.

Table 1
Target words used for the frequency analysis; word frequencies from the Corpus of Contemporary American English (grand total over all registers)

Pair number	Word	Type	Frequency
1	add	+	361,246
	subtract	–	1802
2	addition	+	78,032
	subtraction	–	313
3	plus	+	110,178
	minus	–	14,078
4	more	+	1,051,783
	less	–	435,504
5	most	+	596,854
	least	–	139,502
6	many	+	388,983
	few	–	230,946
7	increase	+	35,247
	decrease	–	4791

3.1.3. Word frequencies

Word frequencies for the words in Table 1 were extracted from the Corpus of Contemporary American English (COCA, Davies, 2010). COCA is balanced for five registers: spoken conversations, magazines, academic language, fiction, and news. Within each of these registers, there are 24 separate text files, one for each year from 1990 to 2012. To analyze frequencies, we used negative binomial regression (Winter & Bürkner, 2021). This type of model was chosen because frequencies are a discrete count variable, and because we expect and want to account for over-dispersion (variance in excess of what would be expected under a regular count model, such as Poisson). Over-dispersion is characteristic of word frequency data (see, e.g., Winter et al., 2018), and hence, negative binomial regression rather than Poisson regression is preferred.

The only fixed effect was "type," which separates addition- and subtraction-related words. As random effects, we included item (paired, e.g., the pair *add* and *subtract* form one item) and by-item varying slopes for the "type" fixed effect. This allows some pairs to have stronger addition bias than others. We also added register and text file as random effects to incorporate as much of the hierarchical structure of the corpus into our multilevel model as possible (cf. Winter & Grice, 2021). Specifically, these random effects were by-register and by-file varying random slopes for the "type" fixed effect.[1]

We used the default priors specified by brms except for the "type" fixed effect, for which we chose the following weakly informative prior: *Normal(0, 2)*. This normally distributed prior is centered at zero, therefore, building in the "mildly skeptical" assumption that most effects are small (McElreath, 2020), and that we should be conservative when data are sparse. The standard deviation represents *how* conservative the model is, with smaller values biasing the model more toward zero. For generalized linear models, weakly informative priors need to be chosen while keeping both the intercept and the nonlinear transformations resulting from the link function in mind (Lemoine, 2019). If we calculated the logged mean for the reference level of this model (= subtract), a weakly informative prior with *SD* = 2 amounts to assuming that 68% of all differences (+/– 1 *SD*) are smaller than 7500 in terms of token frequency, which, as shown in Table 1, is a very conservative assumption. Sensitivity analysis shows that we are able to obtain the main result with an even more conservative prior (*SD* = 1).

3.1.4. Binomial expressions

We created a list of all possible binomial expressions using two linking words (*and*, *or*) from Table 1, that is, *add and subtract, subtract and add, add or subtract, subtract or add, addition and subtraction*, and so on. The corresponding COCA frequencies were collapsed across the *and/or* distinction which is irrelevant for our purposes.[2] The frequencies were analyzed with a mixed negative binomial regression model in the same way that we analyzed word frequencies (including the same item, register, and file random intercepts and slopes, as well as the same weakly informative prior). The main fixed effect was about whether the order was addition-first or subtraction-first. The only difference between the model structure and the word frequency analysis is that we had to add a zero-inflation component to our count model (Bürkner, 2018; Zuur, Ieno, Walker, Saveliev, & Smith, 2009) because many of the binomial expressions we consider are not attested in the corpus, that is, there were more zeros than would be expected under the negative binomial distribution.

It is known that frequency itself is a factor that biases word order in binomial expressions (Benor & Levy, 2006). However, to assess whether there is an addition bias that goes beyond word frequency, one would need to sample a larger number of expressions that show more variation in frequencies. As it stands, our analysis of binomial expressions could, therefore, be an inevitable outcome of the word frequency results we also present in this paper. We nevertheless decided to include this analysis for two reasons: First, it is yet another reflection of the bias in language (regardless of whether it is ultimately caused by frequency or not), and knowing about this is relevant for behavioral studies on addition bias and subtraction neglect, which may use expressions such as *add or subtract*, as was the case in the instructions by Adams et al. (2021). Second, at a deeper level, it is not clear whether the factors of salience and frequency are distinct to begin with, given that salience may be driving frequency, or frequency may be driving salience. For these reasons, our models of the binomial expressions only test the addition versus subtraction difference, not incorporating word frequency as an additional factor.

3.1.5. Emotional valence

Warriner, Kuperman, and Brysbaert (2013) collected emotional valence ratings for 13,000+ English words. Unfortunately, only four of our target words from Table 1 are represented in this dataset. For this reason, we chose to analyze contextual valence, as computed by Snefjella and Kuperman (2016). This contextual valence measure is based on the average emotional valence of the five words immediately preceding and the five words immediately following the target word, which is correlated with the emotional valence ratings of the words themselves ($r = .58$, reported in Snefjella & Kuperman, 2016, p. 139, their Table

3). The contextual valence measure can be thought of as tapping into what corpus linguists call "semantic prosody" (Hunston, 2007; Morley & Partington, 2009; Stewart, 2010), that is, the connotation of a word as is revealed through the surrounding context. When we consider the distributional-semantic closeness to addition-related words in our large-scale analysis of 8000+ English verbs (discussed below), we are not bound by data sparsity and will consider the emotional valence ratings from Warriner et al. (2013).

The contextual valence measure from Snefjella and Kuperman (2016) was available for only four pairs from Table 1. Even though this dataset was sparse, we decided to regress it onto the fixed effect "type" (addition- vs. subtraction-related) in a mixed Bayesian regression with a random effect for pair (random intercept only) and a weakly informative *Normal(0, 0.05)* prior. This prior is adapted to the scale of the contextual valence measure, which has a rather narrow range for this dataset [5.4, 5.7]. Within this range, a normal distribution centered at $SD = 0.05$ embodies the conservative assumption that 68% of all differences are expected to be between [–0.05,+0.05], and 95% of all differences between [–0.1,+0.1].

3.1.6. Distributional semantics (word vectors)

We used pretrained "word2vec" word vectors (Mikolov et al., 2013) from the Wikipedia corpus (Mikolov et al., 2017). We selected word2vec because it is a readily available database of word embeddings based on a large, relatively general-purpose corpus, and because several comparison studies suggest that these word vectors slightly outperform several other off-the-shelf word embeddings in a variety of contexts (Baroni et al., 2014; Beekhuizen et al., 2021; Pereira et al., 2016). One use of word vectors is to look at how much a set of words overlaps with a specific set of probe words (Beekhuizen et al., 2021; Jonauskaite et al., 2021; Lewis & Lupyan, 2020).

Following the discussion in Fischer et al. (2021), our aim was to examine whether verbs of change, as used in the instructions of the behavioral studies conducted by Adams et al. (2021), are semantically biased toward addition. We assessed this by analyzing how such change verbs are similar in semantic space to the verbs in Table 1 that are diagnostic of addition or subtraction, *add/increase* and *subtract/decrease*. We chose the addition- and subtraction-related verbs from Table 1 as diagnostic words since they are most comparable to the verbs we investigate in our analysis of distributional semantics.[3] The verb set we investigate involves synonyms of *to change* and *to improve*, two of the words used by Adams et al. (2021). We selected all synonyms from the Collins dictionary, except for phrasal verbs (e.g., *touch up, polish up*) because word2vec does not include these. The full dataset includes the words *to change* and *to improve*.

Our analysis is based on cosine similarity, a measure of the angle between word vectors. The word vectors are derived from their contextual uses, with the assumption that two words occurring in similar contexts will generally tend to share meaning. Cosine values range from 0 (no similarity in contexts whatsoever) to 1 (complete overlap in contexts). We computed this measure for each verb in the "change" and "improve" set separately, each time with respect to the words *add/increase* and *subtract/decrease*. We then took the average cosine similarity for *subtract/decrease* and subtracted this from the average cosine similarity of *add/increase* to

Table 2
Words considered in the distributional semantic analysis; synonyms of *to change* and *to improve* in the Collins dictionary; our final analysis also included the seed words themselves (*to change* and *to improve*)

Synonyms of *to change*	Synonyms of *to improve*
adjust	*ameliorate*
convert	*amend*
moderate	*augment*
modify	*better*
reform	*embellish*
remodel	*enhance*
reorganize	*mend*
restyle	*upgrade*
revise	
transform	

generate a difference score which represents the degree to which a given word is more similar to addition- than to subtraction-related concepts. In our large-scale analysis reported below, we computed this measure for over 8000 verbs to assess whether words outside the set shown in Table 2 also exhibit a bias toward addition over subtraction.

3.1.7. GPT-3 analysis

Our GPT-3 analyses are based on the “text-davinci-002” model, accessed via the GPT-3 playground GUI (https://beta.openai.com/playground) with the following settings: temperature zero, frequency penalty zero, and presence penalty zero. We enabled “inject start text,” and set “show probabilities” to “full spectrum.” We used GPT-3 in two test cases. The first was based on the sentence frame *The word “…” has a good/bad connotation.*[4] We investigated the cloze probability of the word *good* as opposed to *bad* to see whether a preceding addition-related word made *good* more contextually probable than *bad*, and vice versa for subtraction-related words. The second case for which we used GPT-3 was to prompt the language model with the probe sentence *I suggest we change this by adding/removing*. Similar to our word2vec analysis of “change” and “improve” verbs, we used this sentential context to assess the probability of the word *adding* as opposed to the word *removing* following each one of the verbs from Table 2. This can be thought of as a conceptual replication of our word2vec analysis. We decided to use the word *removing* rather than *subtracting* because the latter is much less frequent, thereby having overall lower cloze probability, and GPT-3’s responses to our prompt revealed that it was “thinking” of the word largely in a mathematical sense. The results we present below would be even stronger for *adding* versus *subtracting*, which is why *adding* versus *removing* is the more conservative comparison. For ease of presentation, we will display probabilities in the figures if they are large enough to be visualized, but we will use log probabilities in our analyses as these give us more granularity in the low probability range.

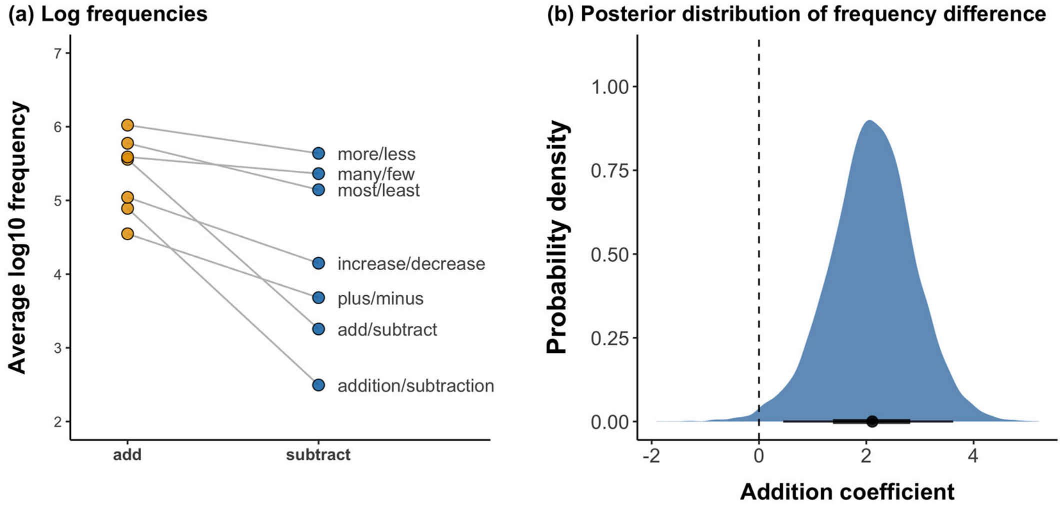

Fig. 1. (a) Log$_{10}$ frequencies by word pairs; (b) Posterior distribution (Kernel density graph) of the type coefficient (log difference of addition to the reference level subtraction) from the mixed negative binomial regression.

4. Results

4.1. Word frequency

As can be seen in Table 1, the total frequency (across all files and registers) was always higher for the addition-related than the subtraction-related target words. The addition-related words were substantially more frequent, as revealed by frequency ratios: from largest to smallest ratios, *addition* was 249.0 times more frequent than *subtraction*; *add* was 200.0 times more frequent than *subtract*; *increase* was 7.8 times more frequent than *decrease*; *plus* was 7.4 times more frequent than *minus; most* was 4.3 more frequent than *least*; *more* was 2.4 times more frequent than *less;* and *many* was 1.7 times more frequent than *few*. Notice that this pattern was without exception: in every single case, the word associated with addition or "more" was more frequent than the corresponding word associated with subtraction or "less." Fig. 1a shows the log$_{10}$ frequencies for each word pair. Pairwise Cohen's d on the log$_{10}$ frequencies shows that the frequency difference was of a large effect size: $d = 1.10$.

The negative binomial regression model generalizing over the random effects of pair, file, and register revealed a strong effect of "type" (addition vs. subtraction) on frequencies, with the estimate (log coefficient = 2.10, $SE = 0.8$) having a 95% credible interval that is clearly above zero [0.45, 3.62]. The posterior distribution of the "type" effect is shown in Fig. 1b. 99% of all posterior samples were above zero. Together with the large Cohen d, this Bayesian analysis suggests that there is substantial evidence for a consistent frequency bias toward addition over subtraction, and this bias in fact characterizes all of our probe words, even when generalizing over the hierarchical structure of the corpus (register, file) via random effects.

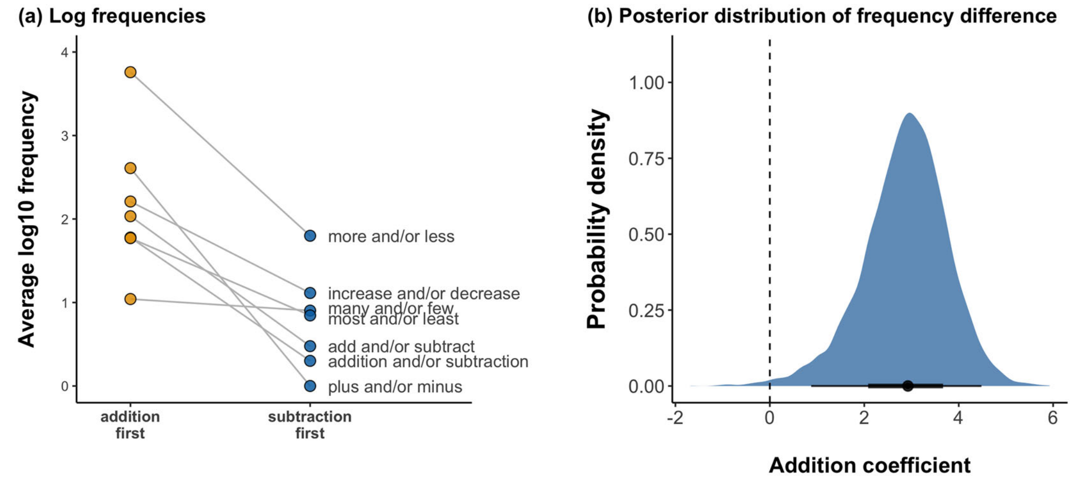

Fig. 2. (a) Log_{10} frequencies by binomial expression; (b) Posterior distribution (Kernel density graph) of the type coefficient (log difference of addition first to the reference level addition second) from the mixed negative binomial regression.

4.2. *Binomial expressions*

English binomial expressions in which the addition-related term comes first (e.g., *add and subtract*) were more frequent than those where the addition-related term comes second (e.g., *subtract and add*), and this was the case for all pairs without exception. This can again be shown by frequency ratios, which were as follows, from largest to smallest: infinite (*plus and/or minus* occurred 430 times, *minus and/or plus* occurred never); 92.5 (*more and/or less* vs. *less and/or more*); 59.0 (*addition and/or subtraction* vs. *subtraction and/or addition*); 53.5 (*add and/or subtract* vs. *subtract and/or add*); 13.4 (*increase and/or decrease* vs. *decrease and/or increase*); 9.7 (*most and/or least* vs. *least and/or most*); and 1.4 (*many and/or few* vs. *few and/or many*). Fig. 2a shows the $\log_{10}$ frequencies for each binomial expression. Pairwise Cohen's d on the log frequencies shows that the frequency difference between addition-first and addition-second has a large effect size: $d = 1.86$.

The negative binomial regression model generalizing over the random effects of pair, file, and register revealed a strong effect of "order" (addition first vs. addition second), with the estimate (log coefficient = 2.87, $SE = 0.9$) having a 95% credible interval that is clearly above zero [0.88, 4.48]. The posterior distribution of the "order" effect is shown in Fig. 2b. 99% of all posterior samples were above zero. Together with the effect size, this Bayesian analysis shows that addition-related words have a strong preference to come first in binomial expressions.

4.3. *Emotional valence*

For all four pairs for which we had contextual valence data from Snefjella and Kuperman (2016), the addition-related words occurred in overall more positive contexts than

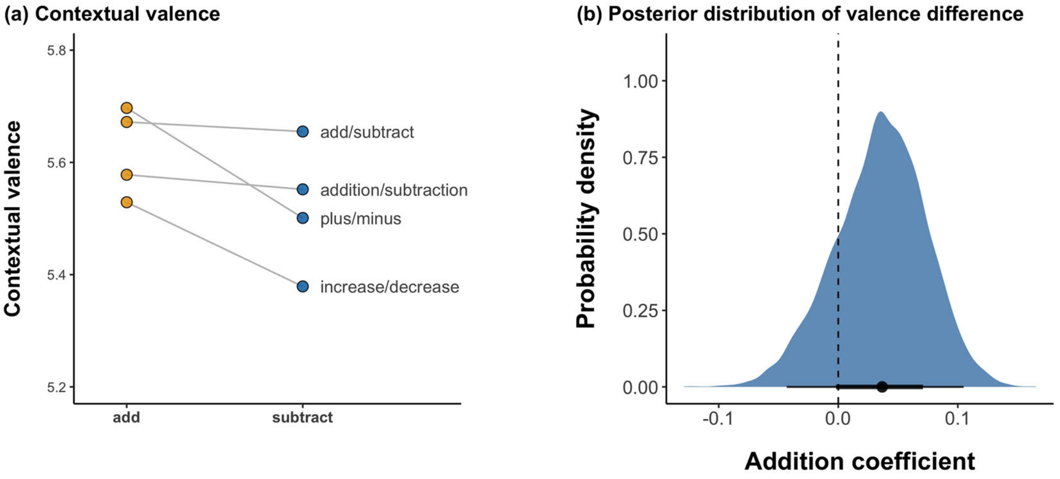

Fig. 3. (a) Contextual valence from Snefjella and Kuperman (2016); (b) Posterior distribution (Kernel density graph) of the type coefficient (difference of addition-related vs. subtraction-related words) from the linear mixed-effects model.

the subtraction-related words, a difference which is associated with a large effect size, $d = 0.94$. The linear mixed effects model with random intercept for pair showed a positive coefficient (with addition-related words being more positive: 0.04, $SE = 0.04$) whose 95% credible interval did, however, substantially overlap with zero [–0.05, 0.11]. Fig. 3 shows the posterior distribution of the coefficient. 81% of all posterior samples are above zero. This means that it is still plausible that the effect could be reversed (19% of all posterior samples), but clearly, more credibility is allocated toward positive coefficients, that is, addition-related words being more positive. More conservative prior choices (narrower standard deviations for the weakly informative prior) make this result even weaker. Such an uncertain result is to be expected given that we are only analyzing four-word pairs. We obtained a numerical trend for contextual valence or what corpus linguists call "semantic prosody," albeit one of large effect size with a direction that is consistent with our "more is better" hypothesis for all tested pairs.

To circumvent the data sparsity issues arising from the lack of overlap with the data from Snefjella and Kuperman (2016), we gave GPT-3 the prompt *The word "…" has a good/bad connotation*, each time with a different addition- or subtraction-related word inside the quotes, for all words shown in Table 1. The reasoning here is that if addition-related words are indeed more positively connoted, the GPT-3 language model should expect the word *good* more in this context than the word *bad*, and vice versa for subtraction-related words. Fig. 4 shows that this is indeed the case: the subtraction-related members of each pair have lower average log probability ($M = –8.70$) in the "bad" context than addition-related words ($M = –7.63$). The effect size for this comparison was large (pairwise Cohen's $d = 1.11$). The picture is reversed, but much less pronounced, in the "bad" context: here, addition-related words are lower in log probability ($M = –9.31$) than subtraction-related words ($M = –8.93$), a small

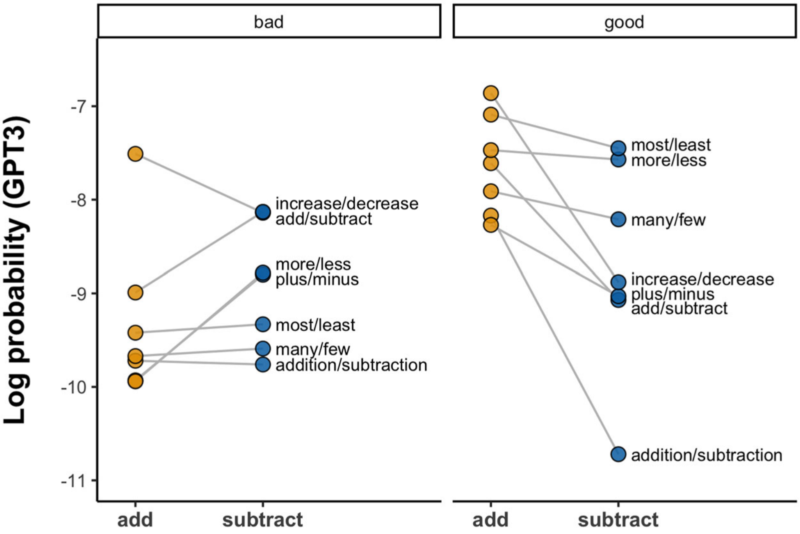

Fig. 4. Log probability of the word *bad* (left facet) or *good* (right facet) for the context *The word "…" has a good/bad connotation* for each word pair.

effect ($d = 0.48$). A Bayesian linear model with predictors for type (addition vs. subtraction), context (good and bad), the type * context interaction, and a random intercept term for word pair provided strong evidence for an interaction, with subtraction-related words being less probable in the good context, compared to addition-related words (–0.37 log probability, $SE = 0.15$). The 95% credible interval of this interaction coefficient did not include zero: [–0.66, –0.07], with 99% of all posterior samples being above zero. These results suggest that GPT-3 expects talk of *good connotation* more so in relation to addition-related words than subtraction-related words.

4.4. Distributional semantics (word vectors)

Fig. 5a shows that without exception, "change" words were semantically more similar in their contextually derived word vectors to the addition-related words than to the subtraction-related ones. The same pattern of results is obtained for the verbs that are synonyms of *to improve* (see Fig. 5b). Cohen's d suggests a small effect size for the "change" words, $d = 0.41$, and a large effect size for the "improve" words, $d = 0.94$.

To assess the reliability of this finding, we fitted an intercepts-only Bayesian regression model to the difference scores (average cosine of addition minus subtraction), analogous to a paired t-test (with an additional *Normal(0, 0.1)* weakly informative prior to make the analysis more conservative).[5] For the "change" words, this analysis revealed a positive estimate of the difference (+0.05), with a small standard error ($SE = 0.01$), and a Bayesian 95% credible interval that excluded zero [0.03, 0.07]. 100% of all posterior samples were above zero. For the "improve" words, the same model structure reveals a similarly robust effect that is stronger in magnitude (+0.08, $SE = 0.01$). Again, 100% of all posterior samples were above zero; the posterior distributions for "change" and "improve" words are shown in Fig. 6.

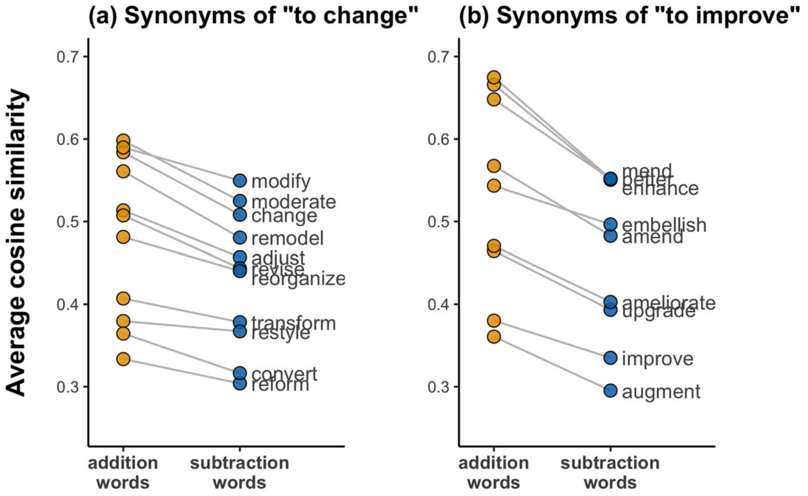

Fig. 5. Average cosine similarities for addition words (*add, increase*) and subtraction words (*subtract, decrease*) for (a) synonyms of *to change* and (b) synonyms of *to improve*; lines represent individual words (paired observations).

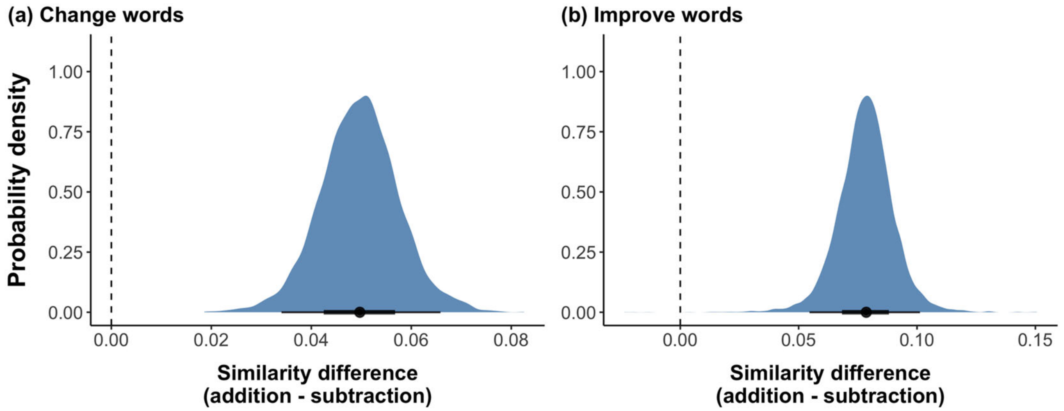

Fig. 6. Posterior distributions (Kernel density graph) of the cosine similarity difference between addition- and subtraction-related words corresponding to our analyses reported in the text; positive values indicate that words are more related to addition than to subtraction; (a) “change” words, (b) “improve” words.

It is possible, however, that this result does not only characterize verbs of change and improvement, but *all* verbs. Could it be that the results discussed so far are inevitable because all words are closer in semantic space to *add/increase* than to *subtract/decrease*? In other words, what is the baseline of cosine similarities and how do the *change/improve* verbs above compare to this baseline? To address this question, we performed a baseline analysis for which we selected 10,000 random verbs from the English Lexicon Project (Balota et al., 2007). Due to imperfect overlap between the English Lexicon Project and the word2vec data, the dataset for analysis resulted in 8,813 verbs. We calculated the addition-over-subtraction bias for each

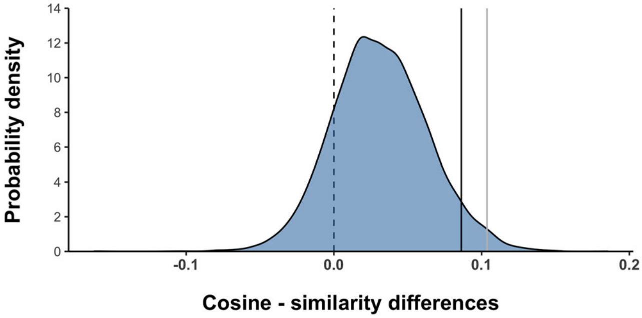

Fig. 7. Kernel density plots of the distribution of cosine similarity differences (closeness to *add/addition* over *subtract/subtraction*) computed for 8813 random verbs; solid black line indicates the average cosine similarity difference of synonyms of *to change*; solid gray line indicates the average cosine similarity difference of synonyms of *to improve*.

of these verbs just as we did before. The resultant distribution of cosine values is shown in Fig. 7, which shows that verbs overall have a bias toward being more related to addition than to subtraction. The synonyms of *to change* were, however, still relatively high in their addition bias, with the average cosine difference of *change* verbs being in the 81th percentile of the random verb distribution; synonyms of *to improve* were in 86th percentile. This analysis suggests that the words *add/increase* occupy a more central position in the overall semantic space (being closer to all verbs), but within this general bias toward addition semantic structure, synonyms of *to change* and *to improve* are even more biased toward addition.

GPT-3 substantiates these results obtained with word2vec. We prompted the language model with *I suggest we "…" this by adding/subtracting*, swapping in different "change" and "improve" verbs, and examining whether this impacts the cloze probability that GPT-3 assigns to *adding* versus *removing*. Fig. 8a shows that GPT-3 does indeed expect the word *adding* much more so after any of the "change" verbs (average log probability = –2.80), compared to the word *removing* (–3.83). The difference between *adding* and *removing* was associated with a large effect size (paired Cohen's $d = 1.31$). A simple intercept-only model fitted on the difference scores between the *adding/removing* case (Bayesian equivalent of a paired *t*-test) estimates the average log probability difference to be 1.01 ($SE = 0.22$), with a 95% credible interval that does not include zero (100% of all posterior samples above zero).

Similar, and in fact stronger, results are obtained for synonyms of the word *to improve*. The probabilities in Fig. 8b demonstrate that GPT-3 expects the word *removing* (average log probability = –4.83) much less after any of the "improve" verbs, compared to *adding* (–2.21). The difference in log probabilities between these two contexts is associated with a large effect size ($d = 3.27$). An intercept-only model fitted onto the difference scores[6] between the two words estimates that *adding* has an estimated log probability that is on average 2.39 higher ($SE = 0.47$), with a 95% credible interval far above zero [1.35, 2.34] (100% of posterior samples above zero). It is noteworthy that a numerically stronger asymmetry between *adding* and *removing* is obtained for synonyms of "improve", given that improving generally has a more positive connotation than changing, which is relatively more neutral. This was also

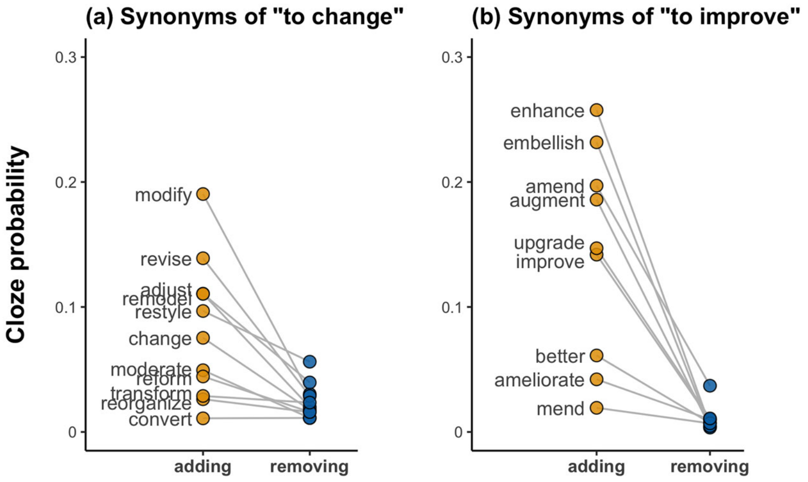

Fig. 8. The probability assigned to the word *adding* or *removing* by GPT-3 in the context *I suggest we change this by adding/removing*; separately for (a) verbs that are synonyms of *to change* and (b) verbs that are synonyms of *to improve*.

apparent in the word2vec analysis, which yielded a stronger addition bias for "improve" than "change" synonyms. The observed numerical trend could be taken as another reflection of the association between addition and positive language.

4.5. *Relating distributional semantics to frequency and valence*

As shown in the above analyses, verbs *in general* appear to be closer in semantic space to *add/increase* than to *subtract/decrease*, albeit this pattern is clearly more pronounced for the "change" and "improve" verbs we consider here. In this section, we use the addition bias score for all 8,813 verbs that overlap between the English Lexicon Project and the word2vec data, which allows us to establish whether the distributional-semantic addition bias is correlated with frequency and emotional valence. Since we are not bound by working with a small dataset for this analysis, we are using the emotional valence ratings by Warriner et al. (2013) here, rather than the contextual valence measure by Snefjella and Kuperman (2016).

We computed our addition bias score (distance in semantic space to *add/increase* minus distance to *subtract/decrease*) for all 8,813 random verbs. With these scores, we can generalize the frequency and valence results beyond the hand-selected target words in Table 1, allowing us to assess whether addition bias influences frequency or valence across a whole swath of verbs. Here, we regressed the addition bias score on emotional valence from Warriner et al. (2013) and the total $\log_{10}$ frequency from COCA (sum across all registers and files) with *Normal(0, 1)* weakly informative priors on coefficient terms. For this analysis, all response and predictor variables were z-scored to facilitate a comparison of effect size. Fig. 9 shows how addition bias can be predicted from (a) frequency and (b) emotional valence.

Fig. 10 shows the posterior distributions of the frequency and valence coefficients from this analysis. There was a positive effect of frequency, with more frequent words also being

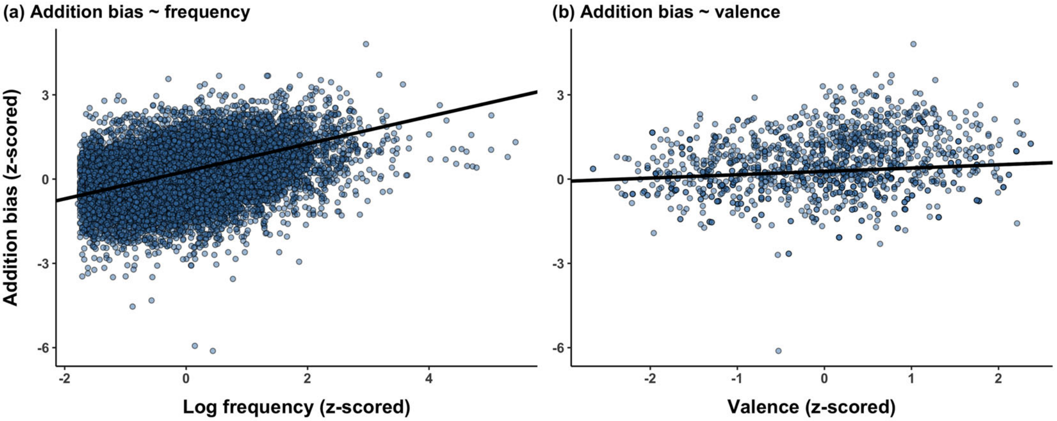

Fig. 9. Addition bias regressed onto (a) frequency and (b) valence, with superimposed lines taken from the simultaneous regression with both frequency and valence as predictors; all variables are z-scored; each datapoint is a word.

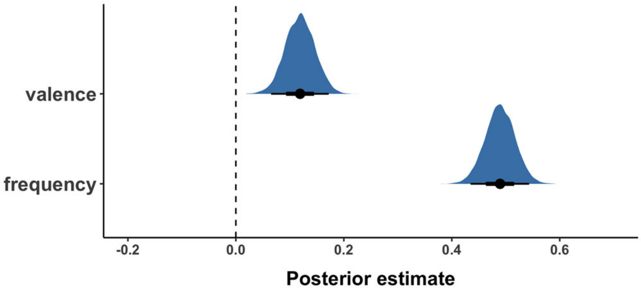

Fig. 10. Posterior distributions for coefficients of a simultaneous regression of addition bias on frequency (COCA) and valence (Warriner et al., 2013).

more addition-biased (0.49, $SE = 0.03$). The 95% credible interval of this coefficient was far above zero: [0.43, 0.54] and 100% of all posterior samples were above zero. The results were much weaker for emotional valence ratings, but there was also a positive effect (0.12, $SE = 0.03$) for which the 95% credible interval did not overlap with zero: [0.07, 0.17]. Although weak, this effect was indicated to be highly reliable, as revealed by the fact that 100% of all posterior samples were above zero.

5. Discussion

Motivated by previous evidence of saliency biases and cultural preoccupations leading to systematic patterns of language production, we examined whether a general addition bias/subtraction neglect observed in behavioral data by Adams et al. (2021) is systematically reflected in language statistics along a number of distinct dimensions. Our results demonstrate that this is indeed the case: the English language reflects a bias toward addition over subtraction in language statistics, as assessed via word frequency statistics, word order preferences in binomial expressions, contextual emotional valence and regular emotional valence, and the distributional semantics of "change" and "improve" verbs. These biases also carry over into the large language model GPT-3. Indeed, the quote that we started this paper with, which GPT-3 generated in response to the prompt *The word "add" is a positive word*, beautifully reflects the extent to which this state-of-the-art natural language processing tool follows the "more is better" heuristic.

It is remarkable just how deep the addition bias runs, which even includes how this paper was written and edited. For example, when writing this paper, we found that it was more natural for us to discuss how addition-related words were over-represented, even though it would have been equally possible to frame things in terms of subtraction-related words being under-represented. And, throughout the previous sections, we followed the preferred order of mentioning addition before subtracting in almost all cases, similar to the results from our analysis of English binomial expressions. This paper also dealt with the addition bias in other ways: during the review process, reviewers overwhelmingly made additive suggestions (additional analysis, additional explanations, etc.), with only one suggestion involving a genuine removal.

We do not claim that our results exhaust the extent to which language reflects a bias toward addition. It is likely that there are other dimensions. For example, given the word frequency result and the general tendency for frequent words to also become shorter over time, words related to addition may also be shorter than words related to subtraction, as is indeed the case for several of the words shown in Table 1. Or, consider the English words *positive* and *negative*: these words have a quantitative sense affiliated with the notions of "more" and "less," but they are also used as evaluative descriptors in line with our emotional valence results.

One obvious limitation of our analyses is our reliance on English data. This is in part motivated because the behavioral studies by Adams et al. (2021) were conducted with English-speaking participants. Another motivation was the ease with which linguistic data could be accessed for English (e.g., emotional valence ratings, language models, etc.). As it is inappropriate to view English as some kind of normative "default case" (cf. Blasi, Henrich, Adamou, Kemmerer, & Majid, 2022; Wierzbicka, 2014), we have to remain agnostic about the extent to which our results generalize to other languages. Future research needs to establish whether other languages exhibit the same linguistic biases. It could even be the case that the extent to which participants from different cultural backgrounds demonstrate subtraction neglect is correlated with the extent to which language reflects this bias, as has indeed been shown for gender biases (Lewis & Lupyan, 2020). It is also possible that cross-linguistic variation in addition bias is modulated by such factors as population size or a country's gross domestic product.

Our findings clearly resonate with Adams et al. (2021), who showed that people are more likely to consider additive rather than subtractive solutions when thinking about tasks or problems. In a commentary to the findings by Adams et al. (2021), Fischer et al. (2021) suggested that the instructions used in their behavioral studies may already be biased toward addition. For example, the instructions mention that participants could *add or subtract*, with no counterbalanced word order (*subtract or add*). Fischer et al. (2021) also performed a simple Latent Semantic Analysis (Landauer & Dumais, 1997) to show that the verbs used in their instructions, such as *to improve* and *to arrange*, were closer in semantic space to addition- than to subtraction-related words. The ideas presented in this commentary gain support from the more extensive analyses presented here, and we show that these ideas extend to other facets of language, such as a word's emotional valence.

From one perspective, our findings can be seen as a conceptual replication of the addition bias identified by Adams et al. (2021), but using language statistics instead of behavioral preferences. From another perspective, our results also point to a chicken-and-egg problem in designing behavioral studies investigating subtraction neglect: given that these studies inevitably have to use language to give instructions to participants, it is possible that the linguistic biases investigated here may influence behavioral preferences. It is possible that language merely reflects an underlying cognitive bias toward addition, but conversely, it is also possible that language itself exerts influence on whether or not participants show subtraction neglect in behavioral tasks. Either way, one needs to be aware of the linguistic biases uncovered here when designing studies on subtraction neglect, and the tools discussed here, such as distributional-semantic vectors, could be used in pursuit of this. For example, to ensure that language exerts minimal bias on behavioral results, researchers could select verbs of change that are least close in semantic space to *add/increase*, for which word2vec or other word embeddings could be used.

The present set of results also connects to the literature on linguistic markedness (Battistella, 1996; Lyons, 1977). For word pairs, such as *odd/even* and *left/right*, the first member is marked (*odd* and *left*), compared to the second (*even* and *right*) (Nuerk, Iversen, & Willmes, 2004). Lyons (1977, as cited in Battistella, 1996, p. 13) distinguished three aspects of lexical marking: formal (e.g., *lioness* is the marked version of *lion* because it is associated with additional morphological material), semantic (as in *young* vs. *old*; *we typically ask how old are you?*, evidencing that *old* is the default case that stands for the whole scale), and distributional (as indicated by occurring in a smaller range of contexts). Markedness is often associated with word frequency, and from this perspective, our results suggest that addition is the unmarked case in English. A number of experiments have shown that linguistic markedness predicts response times in behavioral tasks, such as number classification by parity (reviewed in Fischer & Shaki, 2014; Winter, Matlock, Shaki, & Fischer, 2015). The polarity correspondence hypothesis (Proctor & Cho, 2006) accounts for such congruency effects and biases in behavior by postulating that we mentally align the unmarked poles of stimulus and response dimensions so that concepts such as "addition" and "more" become associated with "good" and "right."

Our results have also applied implications, as do the findings in Adams et al. (2021). In his book-length treatment of subtraction neglect, Klotz (2021) outlines many different real-world situations in which people failed to see subtraction as an option. As is the case with any biases and heuristics, these are not necessarily always detrimental, but we need to be aware of them when making decisions (Kahneman, 2011). Take, for example, the finding that when staff are given opportunities to suggest improvement ideas for a university, only

11% of the suggestions were subtractive in nature (Adams et al., 2021). It is easy to see how such neglect of subtractive solutions, if left unchecked, can lead to rampant bureaucracy, for example, more policies, more forms, and no simplification of procedures. The fact that we documented that even seemingly neutral words such as *to change* already implicitly suggest additive as opposed to subtractive changes in their semantic connotations shows that people need to work extra hard to emphasize the possibility of subtractive solutions in language, so that these will be considered.

Our results also relate to the corpus linguistic literature on the topic of "semantic prosody" in an interesting way. Semantic prosody is a term used by corpus linguists to refer to emotional or affective connotations of words that often cannot readily be intuited when seeing a word in isolation (Hunston, 2007; Morley & Partington, 2009; Sinclair, 1991; Stewart, 2010; Stubbs, 2001; Whitsitt, 2005). For example, the word *to cause* would seem rather neutral without any extra information provided, but this verb is almost invariably used in negative contexts, such as causing death, disease, or war. In the same way, we have shown here that verbs such as *to change* and *to improve* and their synonyms, which do not appear to be particularly related to addition at first sight, are actually biased toward addition when looked at from the perspective of their contextual usage, as revealed through word embeddings and GPT-3. We suggest that corpus linguists interested in semantic prosody should consider incorporating word vectors more closely into their tool kit.

More generally, the present results speak to the utility of word frequency statistics and distributional semantics in uncovering general biases in cultures. As we demonstrated in this paper, the fact that people think more readily about additive solutions across a large range of tasks (Adams et al., 2021) is reflected in language statistics. In line with results from domains, such as perception words (Winter et al., 2018) or environmental terms (Regier et al., 2016), our results lend further support to the idea that language use mimics cognitive biases (see also Louwerse, 2008). Our findings are also broadly in line with the general idea that languages are geared toward communicative efficiency, that is, those aspects of reality that speakers care most about are also those that most prominently feature in language.

Acknowledgments

Bodo Winter was supported by the UKRI Future Leaders Fellowship MR/T040505/1. Andriy Myachykov was supported by the Basic Research Program at the National Research University Higher School of Economics. Martin Fischer was supported by grant DFG-FI-1915/8-1 "Competing heuristics and biases in mental arithmetic."

Data availability statement

All summary data and analysis code used to reproduce the figures in this paper are available under the following Open Science Framework repository: https://osf.io/detuy/

Open Research Badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/detuy/.

Notes

1 The R formula used was: brm (freq ∼ 1 + type + (1 + type | pair) + (1 + type | file) + (1 + type | register), family = negbinomial)

2 The effect size we report in the paper is of a similar magnitude for *and* (Cohen's $d = 1.50$) and *or* ($d = 1.36$). The online repository (https://osf.io/detuy/) contains an additional analysis that incorporates an interaction between addition versus subtraction cases and the *and/or* distinction. As this model is harder to interpret (requiring sum codes to aid main effects in the presence of interactions) and the effect is obtained for both function words separately anyway, we decided to only present the results that disregard the *and/or* distinction for ease of presentation.

3 Our previous analysis (prior to review) was based on the diagnostic word pairs *add/addition* and *subtract/subtraction*. A reviewer suggested that one could also use the entire set of words used in Table 1, which in fact yields the same substantive conclusions (see online Supplementary Materials). We decided that *add/increase* and *subtract/decrease* are the best probe words for this particular analysis, because like the synonyms of *to change* and *to improve*, they are verbs.

4 Comparable (and in some cases even stronger) results are obtained for the sentence frames *The word "…" is a good/negative word* or *The word "…" is a positive/negative word*.

5 Given a normally distributed prior centered at zero, the specific value we chose for the standard deviation ($SD = 0.1$) reflects the assumption that 68% of all differences in cosines would lie between [–0.1, +0.1], and 95% would lie between [–0.2, +0.2]. This value is suitable for the relatively narrow range of cosine values, which, as discussed above, is restricted to 0 and 1.

6 This model used a weakly informative *Normal(0, 1.5)* prior with the standard deviation chosen to slightly bias the model toward zero: for probabilities close to 0.0, a log probability difference of +/– 1.5 entails probabilities between ∼0.0 and ∼0.04 (one standard deviation around zero: 68% of the a priori expected differences), and between ∼0.0 and ∼0.20 for +/ 3 (two standard deviations around zero: 95%).

References

Abdou, M., Kulmizev, A., Hershcovich, D., Frank, S., Pavlick, E., & Søgaard, A. (2021). Can language models encode perceptual structure without grounding? A case study in color (arXiv:2109.06129). arXiv. https://doi.org/10.48550/arXiv.2109.06129

Adams, G. S., Converse, B. A., Hales, A. H., & Klotz, L. E. (2021). People systematically overlook subtractive changes. *Nature*, *592*(7853), 258–261. https://doi.org/10.1038/s41586-021-03380-y

Augustine, A. A., Mehl, M. R., & Larsen, R. J. (2011). A positivity bias in written and spoken English and its moderation by personality and gender. *Social Psychological and Personality Science*, *2*(5), 508–515.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Battistella, E. L. (1996). *The logic of markedness*. Oxford University Press.

Beekhuizen, B., Armstrong, B. C., & Stevenson, S. (2021). Probing lexical ambiguity: Word vectors encode number and relatedness of senses. *Cognitive Science*, *45*(5), e12943.

Benor, S. B., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language*, *82*(2), 233–278.

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, *26*(12), 1153–1170. https://doi.org/10.1016/j.tics.2022.09.015

Boucher, J., & Osgood, C. E. (1969). The Pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, *8*(1), 1–8. https://doi.org/10.1016/S0022-5371(69)80002-2

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*, *10*(1), 395–411.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, *25*(4), 447–464.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*(2), 229.

Firth, J. R. (1957). *Papers in linguistics, 1934–1951*. Oxford University Press.

Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition—From single digits to arithmetic. *Quarterly Journal of Experimental Psychology*, *67*(8), 1461–1483.

Fischer, M. H., Winter, B., Felisatti, A., Myachykov, A., Mende, M. A., & Shaki, S. (2021). More instructions make fewer subtractions. *Frontiers in Psychology*, *12*, 720616.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033.

Heylen, K., Wielfaert, T., Speelman, D., & Geeraerts, D. (2015). Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua*, *157*, 153–172.

Hilpert, M., & Saavedra, D. C. (2020). Using token-based semantic vector spaces for corpus-linguistic analyses: From practical applications to tests of theoretical claims. *Corpus Linguistics and Linguistic Theory*, *16*(2), 393–424.

Houwer, J. D., & Randell, T. (2004). Robust affective priming effects in a conditional pronunciation task: Evidence for the semantic representation of evaluative information. *Cognition and Emotion*, *18*(2), 251–264.

Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics*, *12*(2), 249–268.

Jentzsch, S., Schramowski, P., Rothkopf, C., & Kersting, K. (2019). Semantics derived automatically from language corpora contain human-like moral choices. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*.

Jonauskaite, D., Dael, N., Chèvre, L., Althaus, B., Tremea, A., Charalambides, L., & Mohr, C. (2019). Pink for girls, red for boys, and blue for both genders: Colour preferences in children and adults. *Sex Roles*, *80*(9), 630–642.

Jonauskaite, D., Sutton, A., Cristianini, N., & Mohr, C. (2021). English colour terms carry gender and valence biases: A corpus study using word embeddings. *PLoS One*, *16*(6), e0251559.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kay, M. (2021). tidybayes: Tidy data and geoms for Bayesian models. *R Package Version 3.0.1*.

Klotz, L. (2021). *Subtract: The untapped science of less*. Flatiron Books.

Kloumann, I. M., Danforth, C. M., Harris, K. D., Bliss, C. A., & Dodds, P. S. (2012). Positivity of the English language. *PLoS One*, *7*(1), e29484.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*(3), 1065.

Ladle, R. J., Jepson, P., Correia, R. A., & Malhado, A. C. (2019). A culturomics approach to quantifying the salience of species on the global internet. *People and Nature*, *1*, 524–532.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211.

Lemoine, N. P. (2019). Moving beyond noninformative priors: Why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912–928.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, *20*(1), 1–31.

Lewis, M., & Lupyan, G. (2020). Gender stereotypes are reflected in the distributional structure of 25 languages. *Nature Human Behaviour*, *4*(10), 1021–1028.

Louwerse, M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & Review*, *15*(4), 838–844.

Louwerse, M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302.

Louwerse, M. (2018). Knowing the meaning of a word by the linguistic and perceptual company it keeps. *Topics in Cognitive Science*, *10*(3), 573–589.

Louwerse, M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35*(2), 381–398.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods*, *28*(2), 203–208.

Lyons, J. (1977). *Semantics: 1*. Cambridge University Press.

Malkiel, Y. (1959). Studies in irreversible binomials. *Lingua*, *8*, 113–160.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). CRC Press.

Meier, B. P., Robinson, M. D., & Caven, A. J. (2008). Why a Big Mac is a Good Mac: Associations between affect and size. *Basic and Applied Social Psychology*, *30*(1), 46–55.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., & Orwant, J. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *ArXiv Preprint ArXiv:1301.3781*.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. *ArXiv Preprint ArXiv:1712.09405*.

Morley, J., & Partington, A. (2009). A few frequently asked questions about semantic—Or evaluative—Prosody. *International Journal of Corpus Linguistics*, *14*(2), 139–158.

Nuerk, H.-C., Iversen, W., & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Quarterly Journal of Experimental Psychology*, *57*(5), 835–863.

Pedersen, T. L. (2020). patchwork: The composer of Plots. R package version 1.1.1. Retrieved from https://CRAN.R-project.org/package=patchwork

Pereira, F., Gershman, S., Ritter, S., & Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology*, *33*(3–4), 175–190.

Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLoS One*, *11*(4), e0151138. https://doi.org/10.1371/journal.pone.0151138

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Snefjella, B., & Kuperman, V. (2016). It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition*, *156*, 135–146.

Stewart, D. (2010). *Semantic prosody: A critical evaluation*. Routledge.

Stubbs, M. (2001). *Words and phrases*. Blackwell.

Torchiano, M. (2019). effsize: Efficient effect size computation. https://doi.org/10.5281/zenodo.1480624

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition*, *1*(2), 219–247.

Warriner, A. B., & Kuperman, V. (2015). Affective biases in English are bi-dimensional. *Cognition and Emotion*, *29*(7), 1147–1167.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207.

Whitsitt, S. (2005). A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics*, *10*(3), 283–305.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L. & Hester, J. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Wierzbicka, A. (2014). *Imprisoned in English: The hazards of English as a default language*. Oxford University Press.

Wild, F. (2020). lsa: Latent semantic analysis. *R Package Version 0.73.2*. Retrieved from https://CRAN.R-project.org/package=lsa

Winter, B. (2016). Taste and smell words form an affectively loaded and emotionally flexible part of the English lexicon. *Language, Cognition and Neuroscience*, *31*(8), 975–988.

Winter, B. (2019). *Sensory linguistics: Language, perception, and metaphor*. John Benjamins.

Winter, B., & Bürkner, P.-C. (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, B., & Grice, M. (2021). Independence and generalizability in linguistics. *Linguistics*, *59*(5), 1251–1277.

Winter, B., Matlock, T., Shaki, S., & Fischer, M. H. (2015). Mental number space in three dimensions. *Neuroscience & Biobehavioral Reviews*, *57*, 209–219.

Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, *179*, 213–220.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2), 1–27.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. Springer.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, *10*(4), 954–958.