Stefan Th. Gries
On, or against?, (just) frequency

1 Introduction

Over the past 20 years or so, it seems as if the field of cognitive linguistics has changed quite a bit. While much work during the 1980s and 1990s followed many suggestions by Lakoff, Langacker, Talmy, and others and studied polysemy, metaphor, subjectification, etc., cognitive linguistics now seems to have evolved into a field that is largely construction-based and nearly completely usage-based (see in particular Bybee 2006, 2010, Goldberg 2006). On the whole, I welcome this evolution, but I also sometimes feel that the usage-based part of what cognitive linguistics now is is lacking to a degree that is becoming more and more problematic. The most important problem I see is concerned with the role, maybe since Langacker 1987, of the *f*-word in usage-based linguistics and its relation to Lakoff's (1991) hugely influential cognitive commitment. More specifically, usage-based linguists

- are using the *f*-word too much;
- are using the *f*-word too simplistically; and
- are not considering alternatives to the *f*-word enough,

the *f*-word being, obviously, *frequency*, specifically *token frequency*. In this paper, I want to (i) discuss a few ways in which the heavy reliance on a simplistic notion of frequency is problematic as well as (ii) point to notions that should replace or complement frequency much more in future cognitive-linguistic research if the field wants to do more than merely paying lip service to Lakoff's cognitive commitment.

2 On (token) frequencies

2.1 How (token) frequencies vary

The first major issue with token frequency data (on morphemes, words, (syntactic) constructions, . . .) as they are mostly used in cognitive linguistics is that they are usually based on corpora with too little regard to between- and

Stefan Th. Gries, University of California, Santa Barbara & Justus Liebig University Giessen

within-corpus variability. *Between-corpora variability* refers to the fact that frequencies for the same elements can vary drastically even between corpora that are (supposed to be) representative of similar speakers, registers, and/or time periods. For instance, Schlüter (2005) showed that frequencies of present perfects differ hugely between corpora representing similar as well as different registers of English. Consider also as a much more comprehensive example Roland, Elman, & Dick's (2007) data, which show that the relative frequencies of argument structure and syntactic constructions are, while certainly correlated across corpora, also considerably different across the five corpora they analyzed: Figure 1 shows how the (square roots of the) relative frequencies of their 32 constructions are distributed in the five corpora: the *x*-axis always represents the relative frequency of constructions in the corpus mentioned in the row of the main diagonal, the *y*-axis always represents the relative frequency of the same constructions in the corpus mentioned in the column of the main diagonal, and the dashed line indicates where all points would be if the constructions were equally frequent in both corpora. Even though there is often a sizable correlation (indicated in the diagonally opposite cells), it is clear that there are often multiple outliers.

Even more striking can be a look at *within-corpus variability*, which can be unexpectedly high. For instance, in a follow-up study on Schlüter (2005), Gries (2006) showed that the present perfect's frequency in a single corpus, the British Component of the International Corpus of English (ICE-GB) varies extremely widely, as shown in Figure 2: While the overall mean is around 3%, it comes with a wide degree of variability between the extreme values of 0 and 9.79%.

These are not isolated and/or exaggerating findings: For instance, anyone who has ever worked on first language acquisition of English will probably know that, in the Brown (1973) corpus, Eve's data are quite different from Adam's and Sarah's, or in the Manchester Corpus (Theakston et al. 2001), Ruth's data can look extremely differently from those of the other 11 children. The same is true in second/foreign language acquisition/learning data: Callies (2013) finds a lot of variability in first person pronoun use in MICUSP and CALE, Gablasova, Brezina, & McEnery (2017) show that learner as well as native-speaker data also exhibit a wide spread of frequencies of *I think*, Gries (2018) shows huge individual differences in the use of *quite* in learner and native speaker data, and Wulff & Gries (2019) show that the use of verb-particle constructions can be very different across different learners.

In cognitive linguistics, however, there is much too little work that takes both between- and within-corpus variability seriously. Of the little work that there is on individual differences in cognitive linguistics, that of Dąbrowska and colleagues is probably most relevant. For instance, Street & Dąbrowska (2010) show that

there are "considerable differences in native language attainment" that are correlated with the frequencies of constructions; similarly, Dąbrowska (2018) studies native-speaker attainment of a language and its correlations with both linguistic and non-linguistic predictors. There are other studies that try to consider differences between group and speaker behaviors (e.g., Divjak, Dąbrowska, & Arppe 2016 for another recent example), but on the whole many discussions of frequency effects settle for whole-corpus token frequencies that, at best, approximate a rather crude average of a speaker's input, but no more; cf. Dąbrowska (2016) discussion of Cognitive Linguistics' 'fourth deadly sin' for a similar view of the field's state-of-the-art.
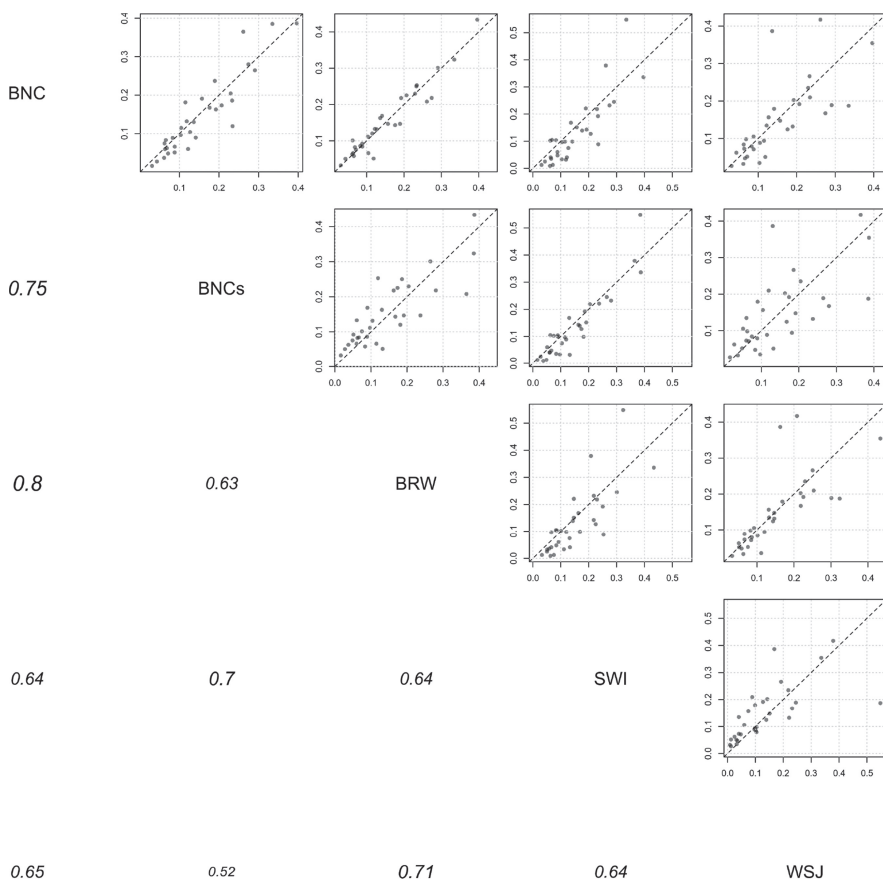
Figure 1: Correlations of frequencies of constructions in the British National Corpus (BNC), the spoken part of the British National Corpus (BNCs), the Brown corpus (BRW), the Switchboard corpus (SWI), and the Wall Street Journal Corpus (WSJ), all data from Roland, Elman, & Dick 2007).
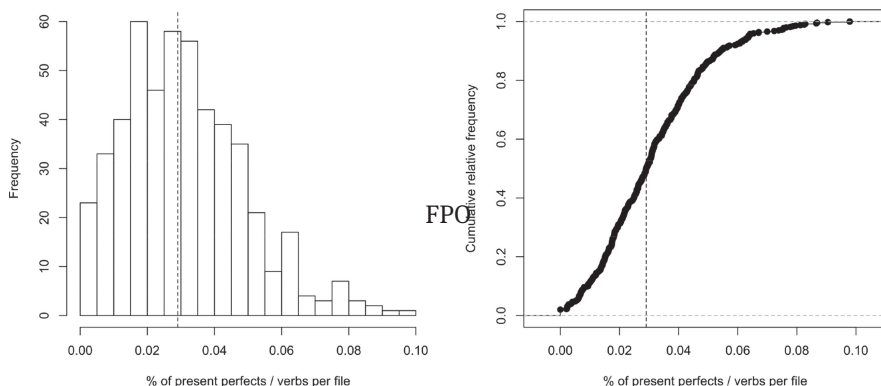
Figure 2: (Relative) Frequencies of present perfects in the 500 files of the ICE-GB.

2.2 How we can deal with this? By increasing the resolution of our corpus studies

The main remedy to this issue is to always increase the resolution on the frequency data: Simple overall corpus frequencies of, say, words or other constructions, are always going to be too much of an imprecise conflation; just like one should never provide a measure of central tendency (e.g., a mean) without a measure of dispersion (e.g., a standard deviation), one should also never provide any overall corpus frequencies without, minimally, a summary indication of their variation in the corpus (e.g. an interquartile range of the frequencies of the construction in question in each part of the corpus, however defined (in terms of meaningful units such as speakers/files, registers, . . .) or, better, an index of the dispersion of the construction in question (Gries 2008). Crucially, this means that a lot of work that is now done only on mostly web-based corpora will need to be done better. Corpora such as COCA, when accessed only in a browser, make it hard to compute such statistics, whereas computing such statistics on corpora whose full text one can access from a hard drive is very straightforward. Cognitive linguistics may need to begin to eschew the convenience of web-access corpora to get higher-quality results.

2.3 What token frequencies correlate with and how much

The next question is how much the notion of frequency actually explains. In much corpus-linguistically informed cognitive or psycholinguistic work, frequency is a significant predictor of acquisition (e.g., in how frequent verbs

drive the recognition of constructional semantics), constructionalization/grammaticalization (e.g., *gonna* or *wanna*), or processing (e.g. in the form of reaction times in lexical decision tasks); the central role that frequency assumes in much psycholinguistic work is neatly summarized by, for instance, Christiansen & Chater (2016:175): "contemporary theories of perception and action have proposed that the cognitive system aims to build a probabilistic model, which captures the statistical structure of the external world". Especially earlier cognitive-linguistic work has also placed a great degree of importance on frequency of occurrence, which has been argued to be an operationalization of the notion of cognitive entrenchment:

> Linguistic structures are more realistically conceived as falling along a continuous scale of entrenchment in cognitive organization. Every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence. (Langacker 1987:59)

However, from a statistical perspective, attempting to support the relevance of frequency with largely monofactorial studies is problematic in a way that is as straightforward as it is often ignored. It is easy to obtain significant correlations between frequency as a predictor and some dependent variable because such a test tests the role of frequency against a null hypothesis that frequency plays no role while controlling for nothing else. However, this kind of testing and similar kinds amount to pretending we do not know anything else already about the phenomenon in question whereas what would really be required is to show that frequency can

– either *complement* what we already know (with controls),
– or *replace* what we already know (but with controls).

This problem is similar to that found in quite a number of studies in learner corpus research (LCR). In many such studies, the frequency of occurrence is actually the response/dependent variable rather than the predictor/independent variable that it is in many cognitive-linguistic/usage-based studies, but the problem is similar in that in many such LCR studies are also monofactorial and consider only one predictor, namely the L1 of the speakers (native speakers vs. different learners). Not only are many of these studies averaging across speakers and even ignoring the sometimes huge percentages of speakers who do not use a certain expression at all (Gries 2018), they also often proceed without regard to any other linguistic/contextual factors that affect the use of the constructions in question.

Thus, many (i) cognitive-linguistic studies that rely on frequency as the main predictor, or even the only one, and many (ii) LCR studies that rely on L1 as the main predictor, or even the only one, share the problem that both leave all the variability that could be explained by many other linguistic, contextual, or psycholinguistic factors one should have included 'up for grabs' by the factors frequency and L1 respectively, leading to anti-conservative overly optimistic estimates of the role of their pet predictors (see Gries 2018 for discussion of this in an LCR context). This, obviously, points to an urgent need for multifactorial explorations of the role of frequency, which in turn requires that we develop accounts of entrenchment or the role of frequency that are more comprehensive in, for instance, including a broader range of predictors (potentially correlated with frequency).

This argument, however raises two questions: (i) is this really necessary? Is there any evidence that calls into question the so widely-attested and seemingly so robust effect of frequency? And (ii) what 'other factors' might those be? Whatever they are, they need to be correlated enough with frequency to explain that frequency has for such a long time been assumed to be so powerful, but at the same time they may be more strongly and even causally related to the phenomena we have so far been explaining with frequency? As it turns out, exploring the former question will begin to address the latter . . .

3 Why token frequency might be less important than is believed

Over the last few years, a variety of studies has indicated that token frequency might not play as much a causal role as has been assumed now for decades. These studies have particularly been concerned with one of the most widely-discussed and robust manifestations of the frequency effect, namely response time latencies or reaction times in lexical decision tasks. This is relevant here because, while there is probably not much prototypical cognitive-linguistic research on such latencies, the kind of frequency effect cognitive linguists have nonetheless been assuming is one that is ultimately grounded in psycholinguistic studies of naming and response latencies. This kind of frequency effects essentially corresponds to a frequency-as-repetition counter along the lines of the above Langacker (1987) quote, or quotes such as "each instance redefines the system, however infinitesimally, maintaining its present state or shifting its probabilities in one direction or the other" (Halliday 1991/2005:67) or "it is usual that each learning event updates a statistical representation of a category independently of other learning events." (Ellis 2002:147). On top of that, the way this frequency

effect is supposed to 'work' is in cognitive linguistics often characterized in terms of (interactive) activation models (also borrowed from psycholinguistics) where more frequent activation of a node is hypothesized to lead to, say, an increase of that node's resting level of activation, which makes it easier for that node to be activated again (after a brief refractory phase, that is) or to an increase in the strength of connections between nodes.

For two to three decades, this view of frequency effects from psycholinguistics has been more or less completely adopted in cognitive linguistics, but in this section, I will discuss a few studies that begin to question the central role of frequency in general or of frequency-as-repetition in particular and then highlight their implications.

3.1 McDonald & Shillcock (2001)

McDonald & Shillcock (2001) discuss a variety of dimensions of lexical variation – frequency of occurrence, concreteness, context availability, age of acquisition, ambiguity – and their correlation with response time latencies. Most importantly, however, they propose a new dimension of lexical variation, one that is correlated with many of the above-mentioned ones, but one that also contains additional information, in particular because, unlike all others, it involves contextual information about words; this, they argue, is necessary because

> [p]sycholinguistic theory has advanced considerably by adopting the convention that lexical representations are discrete entities, and that the meaning of a word can be represented by a simple local representation or by a particular listing of semantic features. In reality it is not possible to provide discrete, necessary, and sufficient representations for the meanings of words; [. . .] It is possible to conclude that the meanings of words are determined by their contexts of usage.　　　　　　　　　(McDonald & Shillcock 2001:300)

The measure they propose is based on co-occurrence information and is called *contextual distinctiveness*; measuring it for a word or a lemma l involves
- retrieving all instances of l within its context;
- computing the relative frequencies of a set of n collocates within a context window around l (e.g., ±5 words); this is the so-called *posterior distribution*, essentially the list of conditional probabilities $p(\text{collocate}|l)$;
- computing the relative frequencies of those n collocates in the corpus in general; this is the so-called *prior distribution*, essentially the list of probabilities $p(\text{collocate})$;
- compute the relative entropy / Kullback-Leibler divergence from the prior to the posterior distribution as in (1).

$$\textit{Contextual Distinctiveness} = \sum_{i=1}^{n} p\left(coll_i \backslash lemma\right) \cdot log_2 \frac{p(coll_i \backslash lemma)}{p(coll_i)} \qquad (1)$$

Thus, contextual distinctiveness "measures the amount of information conveyed by a word about its contexts of use" (p. 303) and is "derived from the distribution of words co-occurring with the word of interest, whereas Word Frequency (WF) is measured independently of this distribution" (p. 307). Contextual distinctiveness is correlated with observed log-transformed word frequency (r=−0.82), but its computation does not involve it *directly* because it is based on co-occurrence percentages; in addition, contextual distinctiveness incorporates prior knowledge (in the form of the probabilities of collocates in the corpus at large).

More important than the theoretical advantages are McDonald & Shillcock's empirical results. In their experiment 1, they find that contextual distinctiveness accounts (marginally significantly) for variance in reaction times in a lexical decision task even when word frequency and length are statistically controlled for (r_{part}=0.2), whereas frequency did not when word length and *CD* were statistically controlled (r_{part}=−0.03). They conclude "[w]ords that appear in relatively constrained (or distinctive) linguistic contexts have high contextual distinctiveness scores and tend to attract longer lexical decision latencies" (McDonald & Shillcock 2011:312). A similar result with regard to partial correlations was then also obtained for data from the lexical decision study carried out by Balota, Cortese, & Pilotti (1999); note in passing that Recchia, Johns, & Jones (2008:271f.) arrive at very similar conclusions:

> lexical processing is optimized for precisely those words that are most likely to be required in any given situation. [. . .] context variability is potentially a more important variable than is frequency in word recognition and memory access.

McDonald & Shillcock then proceed to explore contextual distinctiveness's correlations with the other dimensions of variation and find that it is not derivative from any of the other dimensions but does add something new to the mix since it "has theoretical and empirical advantages over simple word frequency which need to be considered in future research into meaning-based lexical processing behavior" (p. 319).

3.2 Adelman, Brown, & Quesada (2006)

Adelman, Brown, & Quesada (2006) start out from the observation (not often considered in cognitive linguistics at all) that, while psycholinguistic models of

lexical access and reading assume that each encounter of a word allows the word to be processed more quickly later,

> [r]esearch on memory, however, has found that the extent to which the number of repeated exposures to a particular item affects that item's later retrieval depends on the separation of the exposures in time and context (Glenberg, 1976, 1979). Indeed, under some conditions, if neither time nor context changes substantially, there may be no benefit of repetition at all (Verkoeijen, Rikers, & Schmidt, 2004). (p. 814)

From that, they infer that

> [i]f the memory for words that subserves word recognition operates in the same fashion, then the effect of repetitions (i.e., WF) will be diminished or eliminated when these repetitions occur in the same context. Accordingly, the number of contexts in which words are experienced, their contextual diversity (CD), should determine their accessibility and hence response times (RTs) in word naming and lexical decision. A normative measure of a word's CD may be obtained by counting the number of passages (documents) in a corpus that contain that word. (p. 814f.)

It was necessary to quote these passages at length because their argumentation can actually not be left uncommented. This is because, while the empirical studies they report on per se are instructive, the above passage is fraught with some terminological confusion and one critical oversight. First, referring to the "number of contexts in which words are experienced" as *contextual diversity* is not ideal: Just because a "number of contexts" increases does not mean that the diversity of the contexts increases as well. No matter of often *hermetically* is seen in a corpus, the next word will virtually always be *sealed*; no matter how often the expression *was regarded* is seen in a corpus, the next words will virtually always either be *as* or *by*. Yes, given the Zipfian frequency distribution of words in general or in constructionally-defined slots, if one looks at more occurrences of a word, one will ultimately see more different contexts, but this relationship is far from deterministic, and diversity is more usefully operationalized as type frequency in many (cognitive-)linguistic applications.

Second and therefore, the proposal to measure contextual diversity in terms of document frequency is surprising for several reasons. On the one hand, finding a certain word in a variety of documents does not at all guarantee that the actual usage contexts of the word will be different (see *hermetically* and *was regarded* above). On the other hand, Adelman, Brown, & Quesada (2006) do not seem to be aware of the facts that (i) they are suggesting to use what in corpus linguistics has for many decades been referred to as *dispersion*, the degree to which an element is distributed evenly across the parts/documents of a corpus, and that (ii) compared to many of the dispersion measures that have been proposed (see Gries 2008 for

the currently most comprehensive overview), the measure they are proposing – called *range* in corpus linguistics – is probably the crudest one, because it either presupposes that the documents are equally large or it neglects document/corpus part sizes, which will skew the results to some degree.

In their empirical evaluations, they use,

– as dependent variables (responses), reaction time data from six different data bases;
– as independent variables (predictors), log-transformed range and frequency information from three corpora: (i) the Brown corpus (1m words of written American English from the 1960s), (ii) the LSA/TASA corpus of approx. 8.26m words aimed at representing lexical knowledge of 12th-grade high school students, and (iii) the written part of the British National Corpus (90m words of written British English from the 1990s);
– as controls, word lengths, orthographic neighborhood size, rime consistency, number of syllables, and initial phoneme.

Their statistical analysis is based on 18 different regressions on the combinations of six reaction-time data bases and three corpora. They find that, while both word frequency and range add significantly to the explanatory power of regression models already containing the controls, but that "the improvement in prediction was always greater for [range] than for [frequency]" (p. 815).

Adelman, Brown, & Quesada then speculate on whether range is influenced by semantic variables such as ambiguity "as words with multiple meanings should be used in multiple contexts", but I do not consider this too fruitful because, first, as discussed above, occurring in multiple documents/files of a corpus does not guarantee at all that the different occurrences are in actually different contexts and, second, the number of meanings of words is correlated with the frequency of words, another one of Zipf's laws. They do show, however, that range is positively correlated with faster response times "regardless of imageability, concreteness, ambiguity, and other lexical measures" (p. 818) whereas high word frequency is not. The authors conclude that "[l]earning-based models of reading cannot accommodate these results unless they are modified so that learning mechanisms are sensitive to context, not frequency" (p. 822), an interesting conclusion in how it, just like the findings from McDonald & Shillcock (2001), does not really support the frequency-as-repetition counter view that underlies much cognitive-linguistic work using the notion of entrenchment (see Jones, Johns, & Recchia 2012 or Johns, Dye, & Jones 2016 for similar findings regarding the processing of novels words distribution over discourse contexts).

3.3 Gries (2010) and more recent work

Gries (2008) surveyed and critiqued approximately two dozen dispersion measures and adjusted frequencies (frequencies of words that are adjusted downwards if a word's distribution is very uneven). Gries (2010) is a follow-up paper to this publication and is relevant in how it explores more and better dispersion measures than Adelman, Brown, & Quesada. The first part of Gries (2010) is not that relevant to the current discussion because it explores intercorrelations between different dispersion measures to determine to what degree the 16 measures included in the study fall into clusters/components that are internally homogeneous but differ a lot from each other.

The more interesting part is the second, in which Gries computes rank correlations (Kendall's τ) of frequencies with more than two dozen dispersion measures and adjusted frequencies from the 10m words spoken component of the British National Corpus with (i) two of the databases also studied in Adelman, Brown, & Quesada – Spieler & Balota (1997) and Balota & Spieler (1998) – and (ii) reaction time data from Baayen (2008).

The results for the reaction times data from Balota and Spieler show that Gries's own dispersion measure – DP/DP_{norm} – is among those most highly correlated with the reaction time data, but, on the whole, all measures exhibit rather similar correlations: none or no small group is *clearly* superior to the others. This changes with the reaction time data from Baayen (2008), because here there is a clear difference in predictive power between the measures included in the analysis. Gries's *DP* fares well again, but so do some other measures, most of which are measures that – unlike Adelman, Brown, & Quesada's range – correct for differently-sized corpus parts. In fact, frequency is outperformed slightly by range, but range in turn is outperformed considerably by, for instance, the variation coefficient, *D*, and *DP*.

Gries (2019a) is a further application of dispersion measures by (i) using a few more corpora and (ii) measuring correlation in a more flexible way that all studies reported on so far, namely by not using simple linear regression models on the data. As for (i), Gries correlates Balota and Spieler's reaction time data with frequencies and *DP*-values based on the whole BNC, the spoken component of the BNC, the BNC Baby, the BNC Sampler, the Brown corpus, and the British component of the International Corpus of English (ICE-GB). As for (ii), rather than using linear regression modeling (and hoping that log transformations capture all of the non-linearity in the data), he instead uses generalized additive models, i.e. a kind of regression model that can accommodate multiple degrees of curvature in correlation data. The results of separate analyses (for the sake of a rough comparison to Adelman, Brown, & Quesada) are shown in Table 1 and they are

as clear as they can be: For every comparison of *DP* to frequency for each of the corpora and speaker groups, *DP* outperforms frequency, sometimes by more than doubling the amount of deviance accounted for.

Table 1: Percentage of deviance of the reaction time data each explained by a GAM.

	Young speakers		Older speakers	
	DP	Frequency	*DP*	Frequency
BNC	9.26	5.06	17.3	7.57
BNC spoken	8.64	4.26	14.3	5.88
BNC Baby	8.48	4.96	14.9	7.06
BNC Sampler	9.07	5.22	13	6.44
Brown	7.85	4.78	13.2	6.77
ICE-GB	6.1	3.79	9.3	4.78

In other words, despite its ubiquity, frequency is never the best predictor, lending support to the works discussed above.

3.4 Baayen (2010)

The most impressive study on the effect, or lack of effect, of frequency is Baayen (2010), a study explicitly designed to test, among other things, the results of McDonald & Shillcock (2001) discussed above. This study is based on lexical decision latencies for 1042 monomorphemic and monosyllabic words from the English Lexicon Project; here, we will be concerned with two parts of the study, one that determines which predictors (including a token frequency predictor) are correlated with the dependent variable, lexical decision latencies how strongly, and one that determines what the role of of the frequency predictor is vis-à-vis the other predictors. What are these other predictors? In part one of the study they consist of a variety of local and contextual features of words w_{1-n}:

- their contextual diversity and range from above based on measurements from the British National Corpus;
- their textual microcontext based on (i) the type frequency of the immediately preceding word slot, (ii) the entropy H of the w's left syntactic family, (iii) the KL-divergence of the probability distribution of adjectives preceding w and those adjectives' probability distribution in the corpus as a whole, and (iv) the KL-divergence of the probability distribution of prepositions plus

indefinite articles preceding *w* and those prepositions and indefinite articles' probability distribution in the corpus as a whole;
- morphological predictors: (i) the entropy *H* of *w*'s inflectional paradigm, (ii) the noun-verb ratio of the *w*, (iii) *w*'s morphological family size, and (iv) the number of complex words that are synonyms of *w* (according to WordNet);
- word-level predictors: (i) *w*'s neighborhood density, (ii) *w*'s orthographic Levenshtein distance, (iii) *w*'s length (in characters), and (iv) *w*'s letter pair familiarity.

A variety of single-predictor models indicate that the frequency predictor scores the highest R^2-value of all predictors when it comes to accounting for the decision latencies, followed by dispersion (even when only measured by range) and contextual diversity.

However, the more important part of the study for our present purposes is the second one, which explores how predictable the frequency predictor is from the other predictors. This is not only to determine how collinear frequency is with everything else, but also to see what, if any, effect frequency has when all other predictors are residualized out of frequency. As it turns out, all significant predictors account for 91% of the variability of the frequency predictor; one of the strongest predictors is in fact range/dispersion. The most important finding, however, is that, once all other predictors are residualized out of frequency, yielding a bare-bones version of frequency that really only incorporates frequency-as-repetition, that version of frequency
- is still significantly correlated with the original 'regular' frequency predictor, but
- has a very small amount of explanatory power.

In other words, "frequency of occurrence, in the sense of pure repetition, turns out not to be a particularly important predictor" (p. 437). Rather, as further analysis reveals,

> [s]yntactic and morphological family size, dispersion, and syntactic (relative) entropy measures are jointly most predictive, accounting for some 36.7% of the variance. Repetition Frequency does contribute, but [. . .] only 8.8% of the variance is accounted for. This finding replicates the results obtained by (McDonald & Shillcock 2001).

In sum, Baayen's impressive study, while conducted in a psycholinguistic context and focusing on a psycholinguistic concept that, per se, has not been a central issue in cognitive linguistics – lexical access – is one of the earliest comprehensive

studies that should have important implications for cognitive linguistics, some of which will be discussed in what follows.

> so far we have understood neither the nature of frequency itself nor its relation to entrench-
> ment, let alone come up with a convincing way of capturing either one of them or the rela-
> tion between them in quantitative terms.   (Schmid 2010:125)

3.5 Interim conclusions

Cognitive linguistics has for decades now used frequency as the main determinant of cognitive entrenchment, relying on psycholinguistic models largely informed by issues of lexical access and processing (even if cognitive linguists have not always been concerned with lexical access per se themselves). However, given the previous sections, there is strong evidence that at least two other simple factors – contextual distinctiveness and dispersion – outperform cognitive linguistics' pet *explanans* even when it comes to explaining the very kind of phenomenon – lexical access – on which cognitive linguistics have based their reliance on frequencies as an expla-
nation of many many phenomena. And the above studies are not alone; there is a growing body of research that either qualifies the effect of frequency by showing how it interacts with other predictors not often discussed in cognitive linguistics or argues that frequency is generally just less important than is often assumed:

–   Diependaele, Lemhöfer, & Brysbaert (2013) demonstrate that the magnitude of the word frequency effect interacts with speakers' vocabulary size, with weaker effects for those with larger vocabularies (see Preston 1935 for a very early discussion of this interaction);
–   Rayner et al. (2006) find an interaction between frequency and age such that older readers are slower than younger readers, but show a stronger effect of frequency on, e.g., lexical decision and naming latencies;
–   Balota et al.'s (2001) mega study shows that subjective frequency estimates from norming data explain unique variability in lexical decision and naming latencies above and beyond apart from objective corpus frequency (see also Williams & Morris's 2004 results from eye movement latencies). Addition-
ally, Kuperman & Van Dyke (2013) discuss the interaction of frequency and (reading) skill in a variety of processing tasks and show that "corpus-based frequency estimates are not at all reflective of poor readers' true experience with a word, nor can they bring forward the systematically different experi-
ences with common and rare words that readers of varying experience may have", which coincides with Dąbrowska's (2016) deadly sin number 4 again, 'ignoring individual differences'.

In addition to the role of the two factors above and other ones that word frequency at least interacts with, there is also strong evidence that frequency-as-repetition, precisely the notion that is so predominant in cognitive and usage-based linguistics, really does not explain much, at least not in a cognitively/psycholinguistically relevant way. These findings of Baayen's analysis are supported by other similar analyses such as a recent megastudy by Brysbaert et al, which also finds that "when the effects of all other variables are partialed out, there is still a robust word frequency effect (although its impact is diminished to some 5–10% of the variance explained)" (Brysbaert, Mandera, & Keulers 2018:47).

Given all of this, one cannot help but think that the rigor that these studies exhibit and the care with which many psycholinguists try to define and delineate the true causal nature of frequency vs. other factors stand in stark contrast to the often somewhat loose talk about frequency effects in cognitive and usage-based linguistics. It has been quite common and very easy to (i) equate frequency with entrenchment or, more formally, simply operationalize entrenchment using frequency and then (ii) discuss frequency effects with some loose connection to psycholinguistic models whose exact details usually remain unspecified (see Dąbrowska's 2016 deadly sin 2, 'not enough serious hypothesis-testing'). I challenge the reader to read up on cognitive-linguistic studies and compare the number of studies that discuss a frequency effect but remain agnostic about its cognitive/psycholinguistic foundation to those that commit to a testable psycholinguistic model. Thus, while staying at the level of 'general frequency effects' may sometimes be a good-enough work-around to arrive at some first understanding of some data, it falls short of honoring Lakoff's cognitive commitment, which requires much more than the above, namely to be more explicit about the true cognitive (!) underpinnings of frequency and how our knowledge can be informed from other disciplines, here psycholinguistics and corpus linguistics; recall Dąbrowska's (2016) deadly sins number 3, 'not treating the cognitive commitment seriously'.

What needs to be done in order to take the cognitive commitment more seriously? One answer has been the topic of the preceding sections: we need to be more careful in how we deal with frequency in our theory/theories. Here is an admittedly pedantic example of a quotation that highlights at least one problem (but also a first underutilized solution with its mention of recency):

> [t]his seems highly convincing, not least in view of the considerable body of evidence from psycholinguistic experiments suggesting that frequency is one major determinant of the ease and speed of lexical access and retrieval, alongside recency of mention in discourse [. . .]. As speed of access in, and retrieval from, the mental lexicon is the closest behavioural correlate to routinization, this indeed supports the idea that frequency and entrenchment co-vary. (Schmid 2010:115f.)

This quotation includes both causal language (*determinant*) and merely correlational language (*co-vary*) and, thus, makes it harder to infer what mechanism is exactly is envisioned. We need a better understanding of what the variable of frequency does and does not do, we need to determine (better) whether it, or how much of it, is an actual cause or whether it is just correlated with the 'real' causes, we need to get a better idea of what all the things are that frequency is related to how and how much – this also means we may need to consider different versions of frequency to tease apart its components, so to speak – and we need to be clear(er) in what psycholinguistic model we are having in mind or are committing to when we talk about the frequency effect: resting activation levels as a result of repetition, resting activation levels as a result of converging activation from nodes in contexts, profiles of strengths of connections to a node, . . .

However, the other main way in which to take the cognitive commitment more seriously is to explore factors above and beyond frequency. Pertinent suggestions what to look at are already mentioned in places, in fact even in Schmid's quote above: "frequency is one major determinant of the ease and speed of lexical access and retrieval, alongside *recency of mention in discourse*" (my emphasis). Similarly, we find the statement that "[l]earning, memory and perception are all affected by *frequency, recency, and context of usage*" (Ellis, Römer, & O'Donnell 2016:45, my emphasis). Thus, it is no coincidence that the measures discussed above are as powerful and important as they turned out to be: contextual distinctiveness as a measure of how much a word 'warps' the frequency distribution of the collocates (or also more abstract constructions) in its context, and dispersion as one of two manifestations of recency. More specifically, I consider priming the short-term manifestation of recency (because of how priming is related to what a speaker recently processed) and dispersion the long-term manifestation of recency (because, if a corpus is seen as an (obviously) imperfect approximation of what a speaker is exposed to, then dispersion is relatable to what a speaker experienced in the not so recent past). However, despite such statements in theoretical discussions, these factors have received much less attention in empirical cognitive-linguistic studies or even position papers such as those in the special issue of *Cognitive Linguistics* in 2016. This absence is particularly surprising especially for the dispersion component of recency, since one central area in cognitive linguistics has always been the developments of plausible usage-based accounts of language acquisition (i.e. a process fundamentally involving learning, forgetting, and categorization) and processing and previous studies have commented on it explicitly:

> Given a certain number of exposures to a stimulus, or a certain amount of training, learning is always better when exposures or training trials are distributed over several sessions than when they are massed into one session. This finding is extremely robust in many domains of human cognition. (Ambridge et al. 2006:175)

> Schooler & Anderson (1997) also demonstrated that there is a power (i.e., log-log linear) function relating probability of a word occurring in the headline in the NYT on day n to how long it has been since the word previously occurred in that context. The human forgetting curve (Ebbinghaus, 1885) is rational in that it follows this trend.
>
> (Ellis, Römer, & O'Donnell 2016:37f.)

plus recall the above quotes from Adelman, Brown, & Quesada themselves.

It is therefore time that cognitive linguistics at least becomes more aware of how these kinds of studies relativize, or contextualize, our view of frequency and its exact workings and that it explores other dimensions of information. In the following section, I briefly survey a few other dimensions that psycholinguistic work, and at least some cognitive-linguistic work, has found to be important and how they might inform a more comprehensive view of (corpus-based) frequency for cognitive linguistics.

4 What else is there and where does this all (have to) lead?

Above I argued that the notion of frequency is central to cognitive usage-based linguistics, but that such counts do not go far enough. It is useful to reiterate here that my above discussion is critical about the role of frequency as operationalized by the number of corpus attestations of a construction in question and the corresponding view of frequency as a repetition counter – my discussion should not be misunderstood as a blanket attack against frequency data as a whole. This is because, as I have frequently argued elsewhere, of course basically all kinds of corpus-based statistics are ultimately frequency-based: even the computation of contextual distinctiveness or dispersion feature, at some step, the use of frequencies. Thus, my point is specifically that frequencies shouldn't be studied just as frequencies per se (with an accompanying frequency-as-repetition theory), but that they can and should of course also form the input to more sophisticated measures (such as contextual distinctiveness or dispersion). This view is essentially an attempt to take Christiansen & Chater (2016) seriously: if, as they argue, "the cognitive system aims to build a probabilistic model, which captures the statistical structure of the external world," then we should not only go with the simplest/most widely-used kinds of info corpora offer – absolute/relative frequencies of (co-)occurrence – because, are we really assuming that's all the cognitive system does? Of course not, so this section discussed briefly what other kinds of information corpus data have to offer and how they are obtained; I will focus on association/contingency as well as entropy and surprisal.

4.1 Association/contingency

The role of association/contingency can in fact hardly be overstated and Nick Ellis is one of the researchers who has put that notion forward most insightfully on a theoretical level and most forcefully on an empirical level. Ellis (2006) summarized previous work from the psychology of learning as "it [is] contingency, not temporal pairing, that generated conditioned responding in classical conditioning" (p. 10) and that

> human learning is to all intents and purposes perfectly calibrated with normative statistical measures of contingency like r, χ^2 and ΔP [. . .] and that probability theory and statistics provided a firm basis for psychological models that integrate and account for human performance in a wide range of inferential tasks. (Ellis 2006:7)

Thus, it is not just enough to consider how often something happens (e.g. the use of a construction, or the use of a construction with a certain item in one if its slots), but how predictive one (usage) event is of another one, for which one needs to 'normalize' a, say, frequency of co-occurrence against what, in the above example, the word does elsewhere or the construction does elsewhere (see Gries 2012, 2015, 2019b for much discussion for how to do this (best)). The simplest way in which this might be done is to to look at how frequencies of co-occurrence and association measures can return different results. For instance, if one sorts the verbs occurring in the imperative construction in the ICE-GB, then these are the top seven verbs: *be*, *see*, *let*, *have*, *look*, *fold*, and *worry*. This is interesting because several of those seem to intuitively make a lot of sense – *see*, *let*, *look*, *worry* – but (i) *be* and *have* are only in the top seven list because they are very frequent everywhere, but an approach that does not correct for that does not see that their observed frequencies in the imperative are actually less than their expected frequencies (*be* in particular has a high negative log odds ratio to the imperative (nearly −4) and (ii) *fold* is surprising because it only shows up in the imperative in a single file (i.e., it is very underdispersed, see Section 3.2 above) and therefore hardly representative of the imperative anywhere but in books on origami. Thus, we need to adopt more than just frequency: we need association and we need dispersion (all in one tuple, see Gries 2019b) – only then can we get a better resolution on everything that the language learner – L1, L2, FL, . . . – is exposed to and uses in acquisition.

4.2 Entropy and surprisal

Another relevant concept that cognitive linguists need to explore on top of token frequency is that of entropy. For a long time it has been recognized that type fre-

quencies are relevant to (cognitive) linguists in how they are correlated with productivity (e.g. Bybee & Moder 1983, Goldberg 1995: Ch. 5, Bybee & Hopper 2001, Bybee 2010: Section 5.10) and therefore with grammaticalization/constructionalization (e.g. Bybee 2010: Section 6.3) and language learning and acquisition (e.g. Schwartz & Causarano 2007, Endress & Hauser 2011). However, type frequencies, and type-token ratios for that matter, are not all that is relevant because they are not comprehensive enough, especially from a cognitive-linguistic perspective.

For instance, consider Goldberg, Casenhiser, & Sethuraman's (2004) learning experiment: Subjects were exposed to a certain number of tokens (16) instantiated by the same number of types (5). However, the two conditions had different type-token distributions: there was a balanced condition of 4-4-4-2-2 (with an entropy of H=2.25) and a skewed lower-variance condition of 8-2-2-2-2 (H=2). The more skewed distribution was learned significantly better, but this cannot be explained by reference to the type-token ratios of both conditions (because those were identical, namely $^{5}/_{16}$), but it *can* be explained with the distributions' entropies (H=2.25 for the balanced condition and H=2 for the skewed condition). But by now there are a lot of other studies that underscore the relevance entropy has for production/processing:

- Linzen & Jaeger (2015) find that the entropy reduction of potential parse completions is correlated with reading times of sentences involving the DO/SC alternation; e.g., *accept* in *Worf accepted Picard was right* has a lower entropy of possible complementation patterns compared to *forgot* in *Worf forgot Picard was right*, which is reflected in reading speeds.
- Blumenthal-Dramé (2016:500) reports that the entropy of verbs' subcategorization frames correlates with activity in the anterior temporal lobe 200–300 ms after the stimulus.
- Lester & Moscoso del Prado (2017) find that entropies of syntactic distributions affect response times of Ns in isolation and the ordering in coordinate NPs and conclude in a as-construction-grammar-as-it-gets kind of way that "words are finely articulated syntactic entities whose history of use partially determines how efficiently they are produced [. . .] Perhaps words and syntactic structures are much more tightly linked than is typically acknowledged".

Psycholinguistically, the connection between processing and entropy might be explainable in terms of the fan effect, which is "[s]imply put, the more things that are learned about a concept [the more factual associations fan out from the concept], the longer it takes to retrieve any one of those facts" (Radvansky 1999: 198) or within Anderson's ACT-R theory, where the strength of activation S_{ji} between a source of activation j and a fact i is dependent on the log of the fan: "activation [. . .] will decrease as a logarithmic function of the fan associated with

the concept. [. . .] the strengths of associations decrease with fan because the probability of any fact, given the concept, decreases with fan" (Anderson & Reder 1999:188). For the association of a word to constructions, this would mean that the strength of the word's associations will be affected by the number of constructions to which it is connected.

An additional factor that is relevant in this information-theoretic connection is surprisal. Some contemporary learning theories hold that learning is driven by prediction errors: we learn more from the surprise that comes when our predictions are incorrect than when our predictions are confirmed (Rescorla & Wagner 1972). Surprisal is often operationalized as $-\log_2 p$, i.e. the less likely something is, the more we are surprised, and the other way around, where p is typically a conditional probability, e.g. p(verb|construction) or p(function|form) or p(form|function). As Ellis, Römer, & O'Donnell (2016:58) put it, "the surprisal of a word in a sentential context is the probability mass of the analyses not consistent with it". There is now increasing evidence for surprisal-driven language processing and acquisition (Demberg & Keller 2008, Jaeger & Snider 2013, Pickering & Garrod 2013) and it is probably not too far-fetched to consider the possibility that at least some of what surprisal measures will be correlated with another important but often elusive notion in (cognitive) linguistics, salience (Gries 2017:593); also see Gries (2012: Section 5.3) for a discussion of how entropy and surprisal are relevant to discussions of category/construction learning and Jaeger & Weatherholtz (2016) on salience and surprisal for sociolinguistics; however, it seems as if that discussion has only just begun and is fraught with terminological inconsistency (see Zarcone et al. 2016 for a start).

In sum, given both existing empirical findings and possible theoretical explanations for them that are compatible with a cognitive-linguistic framework, it is time that entropy and surprisal be considered and integrated more thoroughly in cognitive linguistics and maybe be afforded a status similarly important as frequency, entrenchment, and other notions.

4.3 Final comments

Let us now conclude and begin with a bit of a 'reminiscent warning' . . . It sometimes seems to me as if (too) much of usage-based linguistics is falling into the same kind of trap much of cognitive linguistics did in the 1990s until Sandra & Rice's seminal article demonstrated the dangers of how liberally notions such as polysemy and semantic networks were used. From a quantitative corpus linguist's point of view, we have now been doing something similar with uncritical assumptions of how frequency 'determines' entrenchment and how frequency is

an important cause for everything . . . How many studies are there that use even one, let alone more, of the more complex kinds of data from above? And how many studies are there that adopt even more of the available methods such as corpus-based prototypicality and semantic-network measures (see Ellis, Römer, & O'Donnell 2016, which should be obligatory literature for any usage-based linguist working with corpora)?

All the above being said, I am *not* trying to say that frequency doesn't do anything or that frequency is not nicely and significantly correlated with many things of interest, so it is certainly very tempting and convenient to use it as an all-purpose tool, the Swiss Army predictor of cognitive/usage-based linguistics. However, if we take seriously Lakoff's cognitive commitment and Ellis's as well as Christiansen & Chater's view of the cognitive system – one that ascribes statistical learning to the cognitive, and thus linguistic, system(s) – then maybe it is time to not just be happy anymore that we found something that 'correlates somewhat well' – we should want more things that actually *cause* and stop pretending that all usage-based linguists need for that is the simplest of statistics, frequency counts. In other words, it is great if usage-based linguists can approximate things well (can we even?) – but it's not great if (i) we already actually know from statistical controls, other studies, . . . that frequency is not necessarily a cause and if (ii) many other factors are available for exploration that do already figure in other cognitive theories (e.g. Anderson's rational theory of learning and memory, e.g. Anderson 1990) or psycholinguistic theories (e.g., expectation-based theories of complexity, e.g. Levy 2008).

One final remark on what that means methodologically. As is clear from the above, the degree of statistical complexity increases once we do not just look at a single frequency list from a whole corpus. And apparently this is scary to many practitioners, as we can see in a recent overview article discussing current challenges, namely when Divjak, Levshina, & Klavan (2016) quote some scholars' views such as "concerns have been raised that the field may be becoming too empirical", "numbers just for numbers' sake", "number-crunching", and "empirical imperialism" . . . These kinds of statements are hugely problematic. It takes a truly interesting view of cognitive linguistics (especially when viewed with Lakoff's cognitive commitment in mind) that condemns the field for becoming too empirical: As if it was unproblematic that the field has been very theoretical for a long time during which unproven theories of various kinds of polysemy networks, metaphorical mappings, and different kinds of construals and scanning abounded . . . (see again also Dąbrowska 2016). As if cognitive linguists were interested in things that are so simple, linear, and accessible to armchair linguistics that one, obviously!, just needed to think about them and maybe eyeball some conveniently small data and/or frequency lists for obvious patterns.

Strangely enough, statements like that sound like they were made by exactly the kind of generative linguists in the 1970s and 1980s, in response to which cognitive linguistics emerged in the first place and, strangely enough, I have yet to see such statements in position papers from psycholinguistics, cognitive science, psychology of learning etc. – how many cognitive science papers do we know that lament that cognitive science is so empirical and is using more and more state-of-the-art statistical methods? If as cognitive linguists we are not just interested in good enough approximations and correlations, but truly interested in causality, then we need advanced statistical modeling of data, which, with its multifaceted kinds of analysis and its proper experimental and statistical controls, will further our understanding of our critical predictors and their causal relations. Avoiding the difficult issues and refusing to engage the line of research Lakoff's cognitive commitment laid out for us nearly 30 years ago is definitely *not* the way forward that the discipline needs to stay vibrant, innovative, but, let's face, also relevant in an ever-changing scientific and theoretical environment . . .

References

Adelman, James S., Gordon D.A. Brown, & José F. Quesada. 2006. Contextual diversity, nor word frequency, determines word-naming and lexical decision times. *Psychological Science* 17(9). 814–423.

Ambridge, Ben, Anna L. Theakston, Elena V.M. Lieven, & Michael Tomasello. 2006. The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development* 21(2). 174–193.

Anderson, John R. 1990. *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, John R. & Lynne M. Reder. 1999. The fact effect: new results and new theories. *Journal of Experimental Psychology: General* 128(2). 186–197.

Baayen, R. Harald. 2008. *Analyzing linguistic data: a practical introduction to statistics with R*. Cambridge: Cambridge University Press.

Baayen, R. Harald. 2010. Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon* 5(3). 436–461.

Balota, David A., Maura Pilotti, & Michael J. Cortese. 2001. Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition* 20(4). 639–647.

Balota, David A. & Daniel H. Spieler. 1998. The utility of item level analyses in model evaluation: a reply to Seidenberg and Plaut. *Psychological Science* 9(3). 238–240.

Brown, Roger. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Brysbaert, Marc, Pawel Mandera, & Emmanual Keulers. 2018. The word frequency effect in word processing: an updated review. *Current Directions in Psychological Science* 27(1). 45–50.

Bybee, Joan. 2006. From usage to grammar: the mind's response to repetition. *Language* 82(4). 711–733.

Bybee, Joan. 2010. *Language, usage, and cognition*. Cambridge: Cambridge University Press.

Bybee, Joan & Carol Lynn Moder. 1983. Morphological classes as natural categories. *Language* 59(2). 251–270.

Bybee, Joan & Paul Hopper (eds). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: John Benjamins.

Christiansen, Morten H. & Nick Chater. 2016. *Creating language: intergrating evolution, acquisition, and processing*. Cambridge, MA: The MIT Press.

Dąbrowska, Ewa. 2016. Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics* 27(4). 479–491.

Dąbrowska, Ewa. 2018. Experience, aptitude and individual differences in native language ultimate attainment. *Cognition* 178. 222–235.

Demberg, Vera & Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2). 193–210.

Diependaele, Kevin, Kristin Lemhöfer, & Marc Brysbaert. 2013. The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology* 66(5). 843–863.

Divjak, Dagmar S., Ewa Dąbrowska, & Antti Arppe. 2016. Machine meets man: evaluating the psychological reality of corpus-based models. *Cognitive Linguistics* 27(1). 1–33.

Divjak, Dagmar S., Natalia Levshina, & Jane Klavan. 2016. Cognitive Linguistics: looking back, looking forward. *Cognitive Linguistics* 27(4). 447–463.

Ellis, Nick C. 2002. Frequency effects in language processing and acquisition. *Studies in Second Language Acquisition* 24(2). 143–188.

Ellis, Nick C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24.

Ellis, Nick C., Ute Römer, & Matthew Brook O'Donnell. 2016. *Usage-based approaches to language acquisition and processing*. New York: Wiley-Blackwell.

Endress, Ansgar D. & Marc D. Hauser. 2011. The influence of type and token frequency on the acquisition of affixation patterns: implications for language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(1). 77–95.

Gablasova, Dana, Vaclav Brezina, & Tony McEnery. 2017. Exploring learner language through corpora: comparing and interpreting corpus frequency information. *Language Learning* 67(S1). 130–154.

Goldberg, Adele E. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago, IL. The University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford. Oxford University Press.

Goldberg, Adele E., Devin M. Casenhiser, & Nitya Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 15(3). 289–316.

Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13(4). 403–437.

Gries, Stefan Th. 2010. Dispersions and adjusted frequencies in corpora: further explorations. In Stefan Th. Gries, Stefanie Wulff, & Mark Davies (eds.), *Corpus linguistic applications: current studies, new directions*, 197–212. Amsterdam: Rodopi.

Gries, Stefan Th. 2012. Frequencies, probabilities, association measures in usage-/exemplar-based linguistics: some necessary clarifications. *Studies in Language* 36(3). 477–510.

Gries, Stefan Th. 2015. More (old and new) misunderstandings of collostructional analysis: on Schmid & Küchenhoff (2013). *Cognitive Linguistics* 26(3). 505–536.

Gries, Stefan Th. 2017. Corpus approaches. In Barbara Dancygier (eds.), *Cambridge Handbook of Cognitive Linguistics*, 590–606. Cambridge: Cambridge University Press.

Gries, Stefan Th. 2018. On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally. *Journal of Second Language Studies* 1(2). 276–308.

Gries, Stefan Th. 2019a. *Ten lectures on corpus-linguistic approaches in cognitive linguistics*. Leiden & Boston: Brill.

Gries, Stefan Th. 2019b. 15 years of collostructions: some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics* 24(3). 385–412.

Halliday, Michael A. K. 2005. *Computational and quantitative studies*. London & New York: Continuum.

Jaeger, T. Florian & Neal E. Snider. 2013. Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127(1). 57–83.

Jaeger, T. Florian & Kodi Weatherholtz. 2016. What the heck Is salience? How predictive language processing contributes to sociolinguistic perception. *Frontiers of Psychology* 7. 1115.

Jones, Michael N. Brendan T. Johns, & Gabriel Recchia. 2012. The role of semantic diversity on lexical organization. *Canadian Journal of Experimental Psychology* 66(2). 115–124.

Johns, Brendan T., Melody Dye, & Michael N. Jones. 2016. The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review* 23(4) 1214–1220.

Kuperman, Victor & Julie A. Van Dyke. 2013. Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance* 39(3). 802–823.

Lakoff, George. 1991. Cognitive versus generative linguistics: How commitments influence results. Language and Communication 11(1–2). 53–62.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar I: Theoretical prerequisites*. Stanford. Stanford University Press.

Lester, Nicholas A. & Fermín Moscoso del Prado Martín. 2017. Syntactic flexibility in the noun: evidence from picture naming. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, 2585–2590.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177.

Linzen, Tal & T. Florian Jaeger. 2015. Uncertainty and expectation in sentence processing: evidence From subcategorization distributions. *Cognitive Science* 40(6). 1382–1411.

McDonald, Scott A. & Richard C. Shillcock. 2001. Rethinking the word frequency effect: the neglected role of distributional information in lexical processing. *Language and Speech* 44(3). 295–323.

Pickering, Martin J., & Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36(4). 329–347.

Preston, Katherine A. 1935. The speed of word perception and relation to reading ability. *The Journal of General Psychology* 13(1). 199–203.

Radvansky, Gabriel A. 1999. The Fan Effect: A tale of two theories. *Journal of Experimental Psychology: General* 128(2). 198–206.

Rayner, Keith, Erik D. Reichle, Michael J. Stroud, Carrick C. Williams, & Alexander Pollatsek. 2006. The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging* 21(3). 448–465.

Recchia, Gabriel, Brendan T. Johns, & Michael N. Jones. 2008. Context repetition effects are dependent on context redundancy. *Proceedings of the Annual Conference of the Cognitive Science Society* 30. 267–272.

Rescorla, Robert A. & Allen R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black & William F. Prokasy (eds.), *Classical conditioning II: Current theory and research*, 64–99. New York: Appleton-Century-Crofts.

Schmid, Hans-Jörg. 2010. Does frequency in the text instantiate entrenchment in the cognitive system? In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*, 101–133. Berlin & Boston: De Gruyter.

Schwartz, Michael & Pei-ni L. Causarano. 2007. The role of frequency in SLA: an analysis of gerunds and infinitives in ESL written discourse. *Arizona Working Papers in SLA & Teaching* 14. 43–57.

Spieler, Daniel H. & David A. Balota. 1997. Bringing computational models of word naming down to the item level. *Psychological Science* 8. 411–416.

Street, James A. & Ewa Dąbrowska. 2010. More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua* 120(8). 2080–2094.

Theakston, Anna L. Elena V.M. Lieven, Julian M. Pine, & Catherine F. Rowland. 2001. The role of performance limitations in the acquisition of Verb-Argument structure: An alternative account. *Journal of Child Language* 28(1). 127–152.

Williams, Rihana & Robin Morris. 2004. Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology* 16 (1–2). 312–339.

Wulff, Stefanie & Stefan Th. Gries. Particle placement in learner English: Measuring effects of context, first language, and individual variation. *Language Learning* 69(4). 873–910.

Zarcone, Alessandra, Marten van Schijndel, Jorrig Vogels, & Vera Demberg. 2016. Salience and attention in surprisal-based accounts of language processing. *Frontiers of Psychology* 7. 844.