

Frequency and entrenchment

Divjak, Dagmar; Caldwell-Harris, Catherine

Document Version
Peer reviewed version

Citation for published version (Harvard):
Divjak, D & Caldwell-Harris, C 2015, Frequency and entrenchment. in D Divjak & E Dabrowska (eds), *Handbook of Cognitive Linguistics*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science, vol. 39, De Gruyter, Berlin.

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Frequency and entrenchment

1. What is frequency?
2. What is entrenchment?
3. Continuing controversies and open questions
4. References

Reference

DIVJAK, D. & C. CALDWELL-HARRIS. 2015. Frequency and entrenchment. In Ewa Dąbrowska & Dagmar Divjak. *Handbook of Cognitive Linguistics*. Berlin: De Gruyter, 53 - 75. [HSK 39].

After half a century of self-imposed exile from the cognitive scene, cognitive linguists are putting language back on stage: language is no longer considered a highly specialized and largely autonomous cognitive module. Instead, cognitive linguists endorse a sophisticated view of learning and memory-based processing. Key to this is the assumption that frequency-sensitive learning results in mental representations optimized for a particular environment. Human beings appear to extract frequency information automatically from their environment (see review in Ellis 2002). Both infants and adults use statistical properties of linguistic input to discover structure, including sound patterns, words and the beginnings of grammar (Saffran et al. 1996). This allows children to learn and adults to refine a probabilistic grammar grounded in our language experience (Diessel 2007; MacWhinney 1998; Saffran 2003).

Whether frequency-sensitive learning really constrains theories of the language faculty remains controversial, however (for an overview of the debate to date, see Lieven 2010; Ambridge and Lieven 2011; Matthews and Krajewski, this volume), and there is a lack of understanding as far as the mechanics are concerned. As recently as 2010, Schmid (2010: 125) concluded his chapter on the relation between frequency in the text and entrenchment in the mind by saying that “so far we have understood neither the nature of frequency itself nor its relation to entrenchment, let alone come up with a convincing way of capturing either one of them or the relation between them in quantitative terms.”

We are less pessimistic. In the current chapter we survey new perspectives on frequency and show how and when frequency-sensitive learning may result in mental representations or memories that vary in robustness and efficiency.¹ Perspectives from both experimental psychology and cognitive linguistics are integrated, with the aim of providing a review that will facilitate future research. We start with the origins of the interest in frequency in cognitive psychology and its interpretation and application in linguistics (Section 1). We then present how the concept of entrenchment has been interpreted in theoretical linguistics, and review the cognitive and neural mechanisms supporting language structures that vary in entrenchment (Section 2). In Section 3 we discuss new directions, controversial issues and open questions.

1. What is frequency?

In experimental psychology, frequency is a practical term that was, and still is, used to capture how frequently a stimulus (such as a word or a phrase) is encountered and processed in the environment.² Within psycholinguistics and cognitive linguistics, frequency most often refers to the number of times a particular chunk of language (such as a phoneme, word, or phrase) occurs in a specified environment. It is typically used in a relative sense, to categorize some stimuli as being more or less prevalent in the environment than other stimuli.

¹ In order to present a coherent narrative in the space available, we have had to omit many relevant papers in the corpus- psycho- and neuro-linguistic traditions. We hope that readers with backgrounds in these areas will understand that these omissions are nothing but consequences of the space limitations imposed, and that readers who are new to these approaches can use the references that we have supplied to find the many interesting studies that we could not cover here. We thank Hans-Jörg Schmid and two further anonymous reviewers of our chapter for their thoughtful comments and suggestions for improvement.

² Because frequency is known to exert a strong influence on processing speed, psycholinguists need to avoid the “confound of frequency” and thus routinely match their experimental items for frequency when comparing reaction times to different categories of words or other language structures.

Frequencies can be obtained in a variety of ways. Some approaches yield subjective results, e.g. asking speakers to estimate frequency of use for a range of language stimuli on a Likert scale from, for example, never encountered to encountered several times a day (Balota et al. 2001). Other approaches yield objective results and rely on counting occurrence of types of stimuli using computer-readable databases or corpora (see also Section 1.4). Historically, most corpora have been drawn from printed text, given the difficulty of transcribing spoken conversations (e.g., Francis and Kucera 1982; Davies 2010), yet many written and spoken corpora now exist in diverse languages (see <http://tiny.cc/corpora> for an overview).³

We first describe the standard ways in which frequency is measured in linguistics. We then provide an overview of frequency effects, i.e. how human beings react differently to higher frequency stimuli compared to lower frequency stimuli. Finally, we draw attention to a range of measures that can help shed light on how frequency effects are rooted in basic brain mechanisms; these measures have been developed within corpus-based and computational approaches but have not (yet) made it to mainstream Cognitive Linguistics.

1.1 Type versus token frequency

Research on frequency in linguistics was given an impetus by the pioneering work of Joan Bybee and collaborators who distinguished between type and token frequencies. The distinction between type and token frequency is important because these two types of frequencies play different roles in the productivity of linguistic structures (Bybee and Thompson 2000).

Token frequency refers to how often a particular form appears in the input, e.g. all instances of the past tense form of *read*, but excluding the present tense form (even though it is spelled identically). Type frequency refers to the number of distinct items that are used in or within the structure of interest “whether it is a word-level construction for inflection or a syntactic construction specifying the relation among words” (Ellis 2002). An example is the number of verbs that create their past-tense by changing an *-ow* form to *-ew*, as in *throw-threw*, *blow-blew*, *grow-grew*.

Token frequency facilitates learning via repetition. The more often a particular token is experienced, the easier it becomes to access and use (Bybee and Hopper 2001). Because it comes with ease of access and use, token frequency can be a conservative force that protects high-frequency structures from analogical leveling.

In contrast to the effects of high type frequency, high token frequency promotes the entrenchment or conservation of irregular forms and idioms; the irregular forms survive because they are high in frequency, which means they are encountered and processed more often (although an irregular form can also survive because it is highly similar to a high frequency item, e.g. *behold*, *forsake*). Type frequency can also guide learners to create a category out of a type (Bybee 1995; Bybee and Hopper 2001). According to Bybee and Thompson (2000), there are three reasons for this:

³ A creative approach to obtaining a large corpus based on spoken language is SUBTL, a large database of frequency norms based on a corpus of subtitles from TV and films (Brysbaert and New 2009). Subjective frequency measures are known to correlate moderately to highly with counts from corpora (Balota et al. 2001; Caldwell-Harris, Berant, Edelman, 2012). Using frequency counts based on a large database of subtitles from TV and films results in higher correlations with processing times than do frequencies from texts Brysbaert and New (2009). This substantiates the intuition that how words occur in dialogue is a more representative measure of their entrenchment than is their frequency of occurrence in written text.

(a) the more lexical items that are heard in a certain position in a construction, the less likely it is that the construction is associated with a particular lexical item and the more likely it is that a general category is formed over the items that occur in that position

(b) the more items the category must cover, the more general are its criterial features and the more likely it is to extend to new items

(c) high type frequency ensures that a construction is used frequently, thus strengthening its representational schema and making it more accessible for further use with new items.

1.2 How can frequency influence processing?

The study of frequency effects has its origin in the seminal psychological research of Cattell (1886). Cattell was the first to demonstrate the word frequency effect, i.e. that higher frequency words are recognized more quickly than lower frequency words. Since the development of the information processing paradigm in psychology in the 1960s-1980s, it has become accepted that frequency is among the most robust predictors of human performance in general (Hasher and Zacks 1984; Howes and Solomon 1951). Human beings are also surprisingly good at providing frequency estimates for a range of language stimuli, suggesting that accumulating frequency information occurs automatically (Hasher and Zacks 1984; Jurafsky 1996; Saffran 2003).

Frequency effects have been found in virtually every subdomain of language that has been studied. Comprehensive reviews of frequency and its effects on first and second language learning, representation and change now exist (Ellis 2002; Diessel 2007; Blumenthal-Drame 2012; Gries and Divjak 2012; Divjak and Gries 2012; Hilpert this volume). Given these reviews, we will focus on providing a taxonomy of types of frequency effects, to set the stage for explaining these effects as the result of frequency-sensitive learning.

1.2.1 Types of frequency effects

It has been most common to study frequency effects using single isolated words, and indeed, the (single) word frequency effect is one of the most robust findings in experimental psychology (Monsell 1991). Frequency effects have also been found for phonemes, morphemes and multi-word expressions, and have been attested for items across the low to high frequency range although less research exists on the former (see Bannard and Matthews 2008, Caldwell-Harris et al. 2012, Snider and Arnon 2012, Divjak under review for recent work on these effects in low frequency structures). Although most of our citations below concern the word frequency effect, note that usage-based linguists propose single-system models and predict frequency effects for all linguistic units: simple and complex, lexical and grammatical.

Frequency effects have been demonstrated for at least five types of behavior:

Faster and easier processing. Using the paradigm of perceptual identification, high frequency words are identified more quickly than low frequency words (Whaley 1978; Monsell 1991). In natural reading using eye-tracking, readers' eye fixations are usually shorter for more frequent words, suggesting greater ease at obtaining the meaning and integrating it with sentence context (Rayner and Duffy 1986).

More accurate processing. Retrieving high frequency items is less subject to error than retrieving low frequency items (Balota et al. 2012; Howes and Solomon 1951; MacKay 1982). When participants are asked to name visually displayed words, a common error is to produce the high-frequency orthographic neighbor of a low frequency target word, as in the case of uttering 'blue' for the target 'blur' (Grainger 1990). Analogous errors are made in spoken word tasks.

Resistance to noise. In visual displays containing ink blots or obscured letters, high frequency words are more accurately detected (McClelland and Rumelhart 1981). In the spoken domain, high frequency words are recognized more accurately when embedded in noise (Pollack, Rubenstein and Decker 1959).

Resilience to brain damage and aging. As semantic dementia progresses from mild to severe, patients have increasing difficulty naming low frequency objects, such as rare animals and items of furniture (Rogers and McClelland 2004). Naming of specific attributes of objects is impaired before naming of more general attributes.

In addition to the behavioral effects of frequency listed above, the neural signatures that accompany language processing vary for high and low frequency stimuli. Event-related potentials (ERPs) measure brain electrical activity that is time-locked to presentation of a word or other linguistic stimulus. A great deal is now known about how wave forms vary for lexical attributes such as word concreteness, word class, semantic ambiguity, and word frequency (Van Petten, 1993). Bigram/trigram frequencies appear to influence the ERP wave form as early as 90 ms after the word is displayed (using single word presentations; Hauk et al. 2006). Lexical (word) frequency has its effect slightly later, at 110 ms post-stimulus onset (Lee and Federmeir 2012). Lexical status, operationalized in these studies as the word/pseudo word distinction, does not influence wave forms until 160 ms, simultaneously with the effects of semantic coherence of a word's morphological family. Researchers have inferred that word frequency influences wave forms earlier than lexical status because word frequency reflects the familiarity with an individual word and its morphologically related forms. In addition, different types of information are believed to be organized in cascades with interactive feedback (Hauk et al. 2006; Rogers and McClelland 2004). We will return to ERP findings later when discussing the role of context in frequency and entrenchment.

1.2.2 Are frequency effects causal?

The frequency with which words occur is strongly correlated with other characteristics (Cutler 1981). Highly frequent words tend to be short in length, concrete rather than abstract, easily imaginable, and have early age-of-acquisition (Whaley 1978). Word frequency also correlates positively with many lexical attributes that have been quantified from corpora, such as orthographic neighborhood density, syntactic family size, noun-verb ratio and number of meanings (Balota et al. 2012; Baayen 2010; Cutler 1981).

Researchers have long suspected that these correlated factors, rather than the extent to which people have been exposed to words, may contribute to the processing

advantage found. To determine how increased usage itself may be responsible for frequency effects, researchers have tried to identify people who could reasonably be expected to have different usage histories. One method has been to compare the lexical processing by persons from different occupations or social groups. In a lexical decision study using nurses, law students and engineers, each group responded more quickly to words relevant to their area of expertise (Gardner et al. 1987). This finding at the word-level was replicated for phrases. Religious Jews have faster processing of religious phrases than secular Jews (Caldwell-Harris et al. 2012). These findings establish that at least part of the frequency effect is due to language users' actual experience with those words and phrases.

1.3 Is it contextual diversity that causes "frequency" effects?

The standard meaning of frequency, and the one we assumed above, is the frequency with which a stimulus is repeated in the environment. This can be called $frequency_{rep}$. Over the last decade, evidence has accumulated that factors which are highly correlated with $frequency_{rep}$ are more strongly correlated with behavioral outcomes than $frequency_{rep}$ itself. One of these factors is the typical context of occurrence of words (Adelman et al. 2006; Brysbaert and New 2009; McDonald and Shillcock 2001).

The discovery of the powerful effect of "contextual diversity" (CD) emerged from data-mining large corpora to extract frequency counts and other values associated with words. Because many words are part of multi-word utterances, researchers sought to understand how much of lexical learning is contextual in nature. McDonald and Shillcock (2001) used principle component analysis over vectors to measure target words' contexts, while Adelman et al. (2006) simply used the number of passages or documents in which words occurred. Even when using this very crude way to operationalize "context", contextual diversity (CD) predicted more variance in lexical decision and naming latencies than did $frequency_{rep}$, suggesting that CD is the psychologically more relevant variable.

Research on explanations for frequency effects turned another corner with Jones and Johns' (2012) claim that what really facilitates lexical processing is semantic diversity. Like Adelman et al. (2006), they counted the number of distinct documents in which a word occurred but defined the similarity of any pair of documents as a function of the proportion of overlapping words in those two documents. A word's semantic distinctiveness was defined as the mean dissimilarity over all of the documents in which it occurred. When used to predict lexical decision and naming times from the Balota et al. (2007) English lexicon database, semantic distinctiveness predicted more variance in response times than word frequency and contextual distinctiveness.

1.4 Contextualized frequency measures

As discussed, psycholinguists have spent decades focusing on word form (token) frequency, and only in the last years have explored alternatives to $frequency_{rep}$ such as contextual diversity. In contrast, among corpus linguists, context has always been a salient issue, and linguists have worked on capturing context in more sophisticated ways. In the section below, we discuss measuring phrase frequency, conditional probabilities, and relational measures from the perspective of corpus linguistics.

Work on lexicography rarely used counts of the occurrence of an individual word form in isolation. This is because words may express different meanings depending on the

context. Classical concordances return a list of usage examples of the item of interest and count its number of occurrences. Words are thus typically examined in their phrasal or sentential context. Indeed, collocations, i.e. words that are regularly used together giving rise to an association, and colligations, where a lexical item is linked to a grammatical one, are important concepts in corpus linguistics. Raw frequencies do not provide a reliable way of distinguishing collocates objectively from frequent non-collocates. The combination of *the* and *review* will be rather frequent due to the frequency of *the*, but *the review* is not a collocation; *peer review*, on the other hand, is. To address this issue collocation scores were calculated that compare expected to observed frequencies to establish whether the observed frequency of co-occurrence is greater than what one would expect to find by chance given the frequencies with which each of the words that form the pair occur. Readers familiar with corpus linguistics will have encountered association measures such as the terms Mutual Information (MI), T-score (Church and Hanks 1990) and Log-likelihood ratio score (or G2, developed by Dunning 1993). The number of association measures available within computational corpus linguistics has grown rapidly over the last decades and we refer to Evert (2005) and Pecina (2009) for exhaustive inventories.

Within linguistics, these mathematically complex measures that capture the strength of association between two items have been perceived to be "so technical that even linguists who had applied them with some success admitted they were not able to see behind the formulas and to interpret the actual linguistic significance" (Schmid 2010: 107). This led Schmid to create conceptually simpler collostructional measures, attraction and reliance (Schmid 2000: 54). Schmid's measures were designed these to capture the interaction between nouns and constructions (rather than the association between two words). They take into consideration the linguistic relation between a so-called node and its collocate, be it another word or a construction, but do not compare observed with expected frequencies. Attraction and reliance were therefore soon supplemented by a set of collostruction techniques (Stefanowitsch and Gries 2003) that pair respect for the relation between a node and its collocate(s) with significance testing. Whether or not statistical significance testing is a desirable property of association measures remains a topic of debate (Schmid and Kuchenhoff 2013, Gries 2013, Divjak under review, Levshina under review).

Corpus linguists have also developed measures of contextual diversity, using the label "dispersion". Dispersion quantifies the homogeneity of the distribution of a word in a corpus (Lyne 1985). Gries (2008, 2010) provides an overview of dispersion measures, including those that penalize words for not occurring uniformly across a corpus. Behavioral data in this area is scarce, but Baayen (2010) shows that dispersion (defined as number of texts in which a word appears) is the second best single predictor of response latencies, after frequency-as-repetition but before contextual diversity (defined as percentage of films containing the word). Although frequency emerges as the best single predictor, frequency of occurrence, in the sense of pure repetition, is not a particularly important predictor in itself, but is instead highly correlated with a number of other factors. It is also interesting to note that dispersion appears to be highly correlated with word frequency ($r=0.82$ reported by McDonald and Shillcock 2001; See also Baayen 2010).

Computational psycholinguists have argued that conditional probabilities (defined as the likelihood to encounter a word given its context, for example)⁴ are more appropriate

⁴ Relative frequencies are conditional probabilities calculated on the basis of one sample only and can be treated as conditional probabilities given a sufficient level of faith in the representativeness of the sample.

than frequencies for explaining language processing in general. Jurafsky (1996) showed that a probabilistic model differs in its predictions from the frequency-based models traditional in psycholinguistics, with true probabilities essential for a cognitive model of sentence processing (cf. Saffran et al. (1996) who showed that infants use transitional probabilities to segment speech and detect word boundaries). The usefulness of probabilities has been well-known within information-theory, where measures such as entropy and surprisal have been developed. Entropy is a measure of the unpredictability of information content: something that is predictable has low entropy, whereas something that is unpredictably has high entropy. In a similar vein, the surprise ratio, also called “suprisal” (Barlow 1990), measures how unexpected a sequence is, given the probabilities of its components.⁵ Suprisal has been used in psycholinguistic models (Hale 2001; Levy 2008; Fernandez Monsalve et al. 2012) and in computational emergentist models (e.g. ADIOS, see Solan et al. 2005).

Contextualized frequency yields better predictions than isolated frequencies, even for low frequency words, and this can be expected: the brain makes use of learned contextual regularities. Seminal studies from the 1970s, such as Biederman et al. (1973), demonstrated already that objects are recognized faster and more accurately when accompanied by contextual information. Although for most research purposes, $frequency_{rep}$ should still be adequate for statistically equating stimuli, it is useful to be aware of alternative measures, since they help address the question of how frequency effects are obtained and are rooted in basic brain mechanisms, a topic addressed later in this chapter.

2. Understanding Entrenchment

Entrenchment was introduced to Cognitive Linguistics as a theoretical construct by Langacker (1987). Langacker used the term entrenchment to explain how linguistic structures is created and shaped through use. A key objective of cognitive linguistics is to determine whether and how the structure of language can result from patterns of usage. It was thus an important step in the foundational writings by cognitive linguists to discuss how linguistic patterns are mentally encoded, and how these representations vary with usage. In this section, we review what linguists mean by entrenchment and connect their theoretical ideas with contemporary views of learning and memory.

2.1 Cognitive linguists' characterizations of entrenchment

In his seminal book, *Foundations of Cognitive Grammar*, Langacker (1987: 59) made the case for 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’s definition of entrenchment focuses on the role of entrenchment for representation, looking at the storage and organization of structures in our mental inventory. Langacker's characterization of entrenchment is noteworthy on two accounts: it

⁵ Hale (2001) showed that the difficulty of a word is proportional to its surprisal (its negative log-probability) in the context within which it appears.

states explicitly that 1) increasing frequency of occurrence deepens entrenchment and that 2) increasing entrenchment can lead to qualitative differences in representation, as when a frequently co-occurring sequence becomes a unit in memory. In other words, it suggests that an increase in frequency deepens entrenchment, and that at a certain point entrenchment may lead to unitization.

Bybee's (2007: 324; cf. also 2007: 10, 279) characterization also emphasizes how repeated use leads to unitization, but she additionally refers to the processing characteristics of automatization and increased fluency or fluidity: "Each token of use of a word or sequence of words strengthens its representation and makes it more easily accessed. In addition, each instance of use further automates and increases the fluency of the sequence, leading to fusion of the units." Important in this second definition is the addition that a number of separate entities can fuse into one larger unit, a phenomenon known as fusion or chunking. Chunk status implies that the unit can be retrieved from mental storage as a whole rather than by accessing the individual component parts and parsing them on the basis of rules or schemas (see also De Smet and Cuyckens 2007: 188).

Blumenthal-Drame (2012: 68f) developed a working definition of entrenchment for her neuroimaging study of multimorphemic words. For this, she drew on concepts of gradedness, fluency, and unitization:⁶

"[h]igher token frequencies in usage will correlate with a gradual increase in ease of processing, more precisely enhanced fluidity in composition or parsing. At some point, this process will lead to a new, holistic representation. After this point, facilitation -- more precisely, ease of retrieval ... -- will still continue to increase as a function of frequency."

Blumenthal-Drame (2012) argued that, crucially, these continuous properties seem to be related to processing, that is to changes in the use of stored entities, rather than the inventory of stored entities, as they imply that the process of fusing separate entities becomes easier and more fluid. She concluded (Blumenthal-Drame 2012:193) that "entrenchment must be seen as a multi-layered phenomenon which is modulated by several stimulus variables and which affects different inter-related yet relatively independent processing dimensions at the same time".

Croft and Cruse (2004: 292) had already stressed the idea that with increasing use, structures continue to accrue representational strength and increase in automaticity, stating that "entrenchment comes in degrees, even beyond the minimum threshold required for independent storage."

From this brief survey, the family resemblance structure among the various characterizations of entrenchment becomes apparent. What these characterizations have in common is the belief that entrenchment refers to a process of strengthening memory representations. This may result in a general reduction in processing effort (automatization), gestalt formation ("unitization" a la Langacker) and/or chunking accompanied by formal reduction ("fusion" a la Bybee).⁷

⁶ There is some terminological proliferation in the entrenchment literature, with several terms pointing in the same direction, i.e. fluency, processing ease, automatization and routinization. We have opted for the term "fluency" to capture both ease in producing and comprehending speech.

⁷ There are also linguists who see entrenchment as a cognitive process to be distinguished from the societal process of conventionalization (Schmid 2010; Mukerjee 2005).

Trying to define entrenchment in theory alone does not seem useful, however, and we now turn to some empirical work on the topic. Within usage-based linguistics proper, most empirical work on entrenchment has been carried out by acquisitionists. A classic question in language acquisition is how children construct grammatical categories and rules when adults rarely correct childrens' grammatical errors, an issue related to poverty of the stimulus arguments (Pullum and Scholz 2002). Attending to frequently occurring constructions can mitigate the lack of negative evidence (Braine and Brooks 1995). These authors propose what they call the "entrenchment hypothesis": repeated presentations of a verb in particular constructions (e.g., *The rabbit disappeared*) cause the child to probabilistically infer that the verb cannot be used in non-attested constructions (e.g., **The magician disappeared the rabbit*). Learning from positive evidence will create verb-argument structures which have a strength proportional to how often a verb has been heard with that argument structure (this line of inquiry is taken further by work on statistical pre-emption, see Goldberg 2011; Boyd and Goldberg 2011; Casenhiser and Bencini, this volume).

An implication of this view is that when an argument structure has been learned to a stable level of entrenchment, it will pre-empt alternatives, unless they have been independently witnessed. A second implication is that overgeneralizations will be less common, and will subjectively feel less acceptable, for high frequency verbs than for semantically-matched lower frequency verbs. For example, **The magician vanished the rabbit* feels slightly more acceptable than **The magician disappeared the rabbit*, since this inference from absence is stronger for the higher-frequency verb *disappeared*. Ambridge (2013) confirmed that children were more accepting of low frequency verbs being used in novel high frequency constructions, than of high frequency verbs being used in alternative constructions. For alternating ones, such as the dative and locative constructions, the effects were less pervasive (see Ambridge and Lieven 2011: 252-254). This leaves open the question of how speakers deal with newly witnessed or rarely attested alternatives: since they have been independently witnessed they should no longer be pre-empted on a strict entrenchment account.

Like other researchers, Braine and Brooks (1995: 368) did not take a stance on the precise quantitative relation between representational strength and frequency of usage. They merely note that with age there appears to be an increase in flexibility in switching between sentence constructions to meet conversational demands (e.g. to have particular arguments as subject or as object). Our contribution here will therefore be to draw insights about learning and memory from cognitive psychology, so that cognitive psychology can underpin Cognitive Linguistics.

2.2 Entrenchment: what learning does to the brain

To be maximally helpful to linguists who want to draw on insights from cognitive science and learning, we suggest a working definition of the relation between frequency and entrenchment. Frequency facilitates language processing because the available mental representations have been shaped by frequency-sensitive learning. As such, they are prepared to process stimuli that vary widely in their probability of occurrence in the environment (Elman 1993; Saffran et al. 1996). From a cognitive science perspective, mental representations can be considered stable attractors in the brain's dynamic neural networks

(MacWhinney 1998). These dynamic patterns vary along a continuum of strength of representation.

2.2.1 The neurocognitive basis of entrenchment

How are “strong” representations (or “deep attractors”) mentally represented differently from weaker representations? There are several ways to conceive of representational strength, as has been done via modeling in artificial neural networks (Rogers and McClelland 2004). Strength of representations can correspond to heavily weighted connections from some input features to processing units inside the networks' architecture. There can also be a large numbers of connections, and more redundant connections. Weighted connections between processing units are functionally akin to neurons' dendrites and axons. Specific links between processing units that frequently match inputs to their expected outputs are strengthened, inspired by the Hebbian learning principle (Hebb 1949) in neuroscience that “neurons that fire together wire together”.

It may seem odd to equate entrenched linguistic forms with something as prosaic as “memory”. But entrenched forms must be memories (Bar 2011; Daelemans and Van den Bosch 2005). Memories capture information that has been encoded and can influence future processing; there is no requirement for memories to be conscious or to be recallable. This is clear from the classic distinction between declarative and procedural memories, also termed explicit and implicit memory. Declarative memories are those for which we have conscious recognition, including episodic memories. For language stimuli, we may be able to consciously recall autobiographical episodes when a specific word or phrase was used. Or we can have recognition memory –and be able to reliably confirm that a phrase such as “about which” is familiar and we have likely used it thousands of times. We can also confirm that the phrase “which about” is not familiar and indeed we may never have used it; it is highly unlikely to exist as an entrenched unit (Caldwell-Harris et al. 2012).

2.2.2 Is there a threshold number of occurrences required for entrenchment?

The cognitive science perspective provides a framework for thinking about some of the outstanding questions in the relationship between frequency and entrenchment. It is a common view, but controversial, that a stimulus sequence needs to be encountered a certain number of times before it becomes unitized (i.e., encoded as such in memory). According to this view, once complex stimuli are encoded as units, their mental representations grow in strength as a function of experience. This common view lacks empirical support (Gurevich, Johnson and Goldberg 2012). Researchers have not been able to find evidence of what might be a frequency threshold for multimorphemic or multiword utterances. Alegre and Gordon (1999) have proposed a threshold of 6 occurrences per million words for inflected forms, but frequency effects have been observed well below that threshold (Baayen et al. 1997; Baayen et al. 2007; Blumenthal-Drame 2012; Arnon and Snider 2010; Caldwell-Harris et al. 2012, Divjak under review), and are found for all frequency ranges for morphologically simple controls (Alegre & Gordon 1999).

A second counterargument is logical. If a single exposure is below the threshold where a counter begins accruing evidence, then the counter of exposures remains set to 0, and logically no experience can accrue (Gurevich et al. 2012). It may be more fruitful to assume that evidence accrues from the first exposure, but that speakers cannot formulate

reliable hypotheses until sufficient evidence has accumulated: Divjak (under review) finds frequency effects for rare lexico-syntactic combinations in Polish (<.66 pmw) but shows that these effects are driven by words that themselves occur at least 6 times pmw. Erker and Guy (2012) propose to think of frequency as a gate-keeper or potentiator: some constraints on subject personal pronoun use in Spanish are activated or amplified by high frequency. This is expected on a probabilistic approach to language, and can also be explained by what we know from memory research, in particular from research on how information is transferred from immediate working memory to long term memory.

2.2.3 The role of procedural and declarative memory systems

Memory for specific episodes is believed to be part of the declarative memory system, mediated by the hippocampus and medial temporal structures (Cohen and Squire 1980). The declarative memory system performs one-trial learning, but such information is subject to rapid decay. Recurring events are learned via the procedural system, mediated by neocortical structures (Gupta 2012). Here, slow learning allows information to be incrementally integrated into long term memory structures, where they have rich associations with many patterns, facilitating generalization and abstraction.

Connectionist models have been used to describe how human languages draw on both the procedural and declarative systems for learning (Gupta 2012; Rogers and McClelland 2004). The procedural system is most efficient at encoding systematic mappings using distributed representations. In distributed representations, multiple patterns are stored across the same set of processing units, allowing for extraction of regularities. Novel patterns can be rapidly learned via minor changes to the weighted connections in the network, but these minor changes will typically be overwritten again as soon as new patterns come in.

Learning unique arbitrary mappings, such as the link between word forms and meanings, can be done if sparse or localist representations are used, since the patterns won't interfere with each other. It has been proposed that hippocampal structures use sparse representational structures to implement arbitrary associations, including episodic and short-term memories (Rogers and McClelland 2012). Arbitrary associations can be permanently learned only with considerable exposure/training. Theorists propose that with continued rehearsal and learning, these associations are gradually displaced from the fast-learning hippocampal system and integrated into the neocortical procedural system.

Learning lexical items, morphological patterns and syntactic constructions is complex and relies on the integration of these two brain systems (see Gupta 2012 for a review). Learning a new morphological variant can usually be handled by the procedural system because it involves minor adjustments to established sound-to-motor patterns. Novel mappings, such as learning to pronounce a foreign word or learning someone's name, require creating new pathways between inputs and outputs, and thus may be initially stored as part of episodic memory. If that novel information is never encountered again, the weighted connections that represent it will be overwritten as new patterns are encountered. But if that stimulus is repeatedly encountered, each exposure provides another training trial in which it can be integrated into long-term memory structures in the neocortex.

2.2.4 Encoding in context

Appreciation is growing that language processing has more in common with memory retrieval than has previously been assumed (Adelman et al., 2006; see also the computational linguistic project called memory based learning (MBL), Daelemans and Bosch 2005).

The brain mechanisms that underlie entrenchment specify a major role for repeated experience, whether it is overt experience in the environment, or mental rehearsal during silent rumination. The best recall is for material that has been encountered at varying times and locations (i.e., in separated contexts).⁸ To explain why words with high contextual diversity are recognized more quickly, Adelman et al (2006) turned to research on the advantage of spaced exposures for long-lasting learning (Anderson and Schooler 1991). Exposures that are widely spaced in time and occur in different contexts have the strongest impact on learning. The reason is that repeated stimuli that re-occur immediately may be processed as if they were a single episode, because of the phenomenon of repetition suppression (Grill-Spector et al. 2006). When a word (or other stimulus) is presented twice in rapid succession, the second occurrence is considered 'primed' - it is more easily processed compared to following an unrelated stimulus (Lee and Federmeir 2012). But this ease-of-recognition brings with it reduced neural activation. This repetition suppression plausibly results in less opportunity for strengthening connections, meaning that less learning (and less entrenchment) occurs for items that are encountered repeatedly in a short period of time. Not surprisingly, people have the poorest recall for "massed practice", meaning training on items that are encountered within a defined time period, or in a single, predictable context, as is typical of classroom academic learning. High frequency_{rep} thus does not in and of itself ensure integration into long term memory structures.

Another relevant line of thought comes from the perspective of "rational analysis of memory", which posits that it is adaptive from an evolutionary perspective to only encode items which are likely to reoccur in the future (Anderson and Schooler 1991). Indeed, a view from evolutionary and cognitive psychology is that the purpose of memory is not to remember past events, but to have mental resources to guide future action (Bar 2011). The greater the diversity of environments in which something has occurred in the past, the more probable is it to reoccur in the future. Simple frequency_{rep} therefore strengthens an item's representation less than if the item was experienced in a different context.

Effects of contextual diversity appear to arise naturally in a learning model that includes context. Baayen (2010) found that contextual diversity is an emergent property of a computational model originally developed to explain morphological processing, the naive discriminative reader (NDR; see also Baayen & Ramscar, this volume). In the NDR model lexical meanings are learned from contextually rich input.⁹ These are letter bigrams and trigrams drawn from a window of four words rather than from words in isolation. The activation of a meaning on a given trial is obtained by summing the weights from the active letters and letter pairs to that meaning. The NDR model correctly predicted a range of

⁸ This is the same finding as from the educational literature, where cramming for a test yields less enduring learning than do spaced study periods (Carpenter, Cepeda, Rohrer et al., 2012).

⁹ The NDR model shares some features with connectionist models, using an error-driving learning algorithm to map from inputs (representations of letters) to outputs (representations of meanings). It differs from connectionist models by using straightforward symbolic representations for letters, letter pairs and meanings. It only uses one forward pass of activation, with weights set on links having been computed from corpus-derived co-occurrence matrices.

morphological phenomena and showed contextual diversity effects. The contextual diversity accounted for substantially more variance in word recognition efficiency than did word frequency. Another success of the model was that it also predicted phrase frequency effects (Baayen and Hendrix 2011), which are known to be quite robust (Arnon and Snider 2010). Other computational models, such as the memory based learning approach (Daelemans and Bosch 2005) have likewise reported that token frequencies of linguistic patterns do not enhance classification accuracy.

3. Continuing controversies and open questions

In this final section, we highlight a few of the controversies and open questions concerning frequency and entrenchment within Cognitive Linguistics. In our view, entrenchment is best thought of as the procedure that gives rise to mental representations through frequency sensitive learning. These mental representations are effectively memories, and thus concepts from current work on memory apply. Taking a broader cognitive science perspective also has the advantage of offering new points of view for two commonly asked questions about the relationship between frequency and entrenchment.

3.1 What can be entrenched?

An frequently asked question is: what can be entrenched? Single words, complex phrases, lexical items, abstract schemas? If entrenched expressions are mental representations of language forms which are either implicit or explicit memories, then, yes, all of these can be entrenched. The more difficult question is whether entrenchment necessarily implies chunking and chunk storage

It has been common practice to view frequency effects as proof of the existence of mental representations. If frequency effects were found for a specific morpheme sequence, then researchers felt justified in viewing that morpheme sequence to be mentally represented as a discrete unit. For example, Blumenthal-Drame (2012:193) concluded from her study of the processing of multimorphemic words that “[...] the effects of token frequency at different levels of language description attest to the necessity of positing full storage of tokens, irrespective of whether they are complex or simple” (cf. also Bannard and Matthews 2008; Arnon and Snider 2010).

Recent computational modeling casts doubts on the wisdom of these assumptions. Baayen’s (2011) naive discriminative learner model contained no representations corresponding to whole words or phrases, only letter unigrams and bigrams (see also Baayen & Ramscar, this volume). The model nevertheless showed frequency effects for multi-word units. Based on this demonstration, Baayen (2011) argued that specific morpheme sequences (including multiword expressions) show frequency effects: the model develops its own representations that are frequency sensitive, as a by-product of learning form-to-meaning mappings that vary in frequency.

3.2 Can we resolve the tension between storage and computation?

Another take on this problem comes from the discussion about the relationship between storage and computation. It continues to be debated whether frequency effects

are observed because a frequent multimorphemic word or multiword expression is stored as a unit or whether its pieces are more rapidly assembled.

Blumenthal-Drame (2012: 187) argued that “[...] highly independent representations will be holistically retrieved rather than analytically processed.” Tremblay et al. (2011: 595) provided evidence for holistic storage but noted at the same time that behavioral research may not be able to distinguish between holistic retrieval and speeded online computation. Other researchers have suggested that the tension between storage and computation is unnecessary. Shaoul (2012: 171) proposed that “this graded effect of probability (...) is a side-effect of the emergent nature of n-gram processing”.

In other words, the neural patterns which mediate language processing contain probabilistic expectations of how patterns will be completed. Any given syllable encountered in a speech stream activates expectations for a subset of all possible syllables based on prior processing (Elman 1993; Baayen and Hendrix 2011). Expectations are activated quickly and effortlessly, as if the predicted sequence was stored separately as a ready-made unit (see Baayen and Ramscar, this volume). This view of expectation generation and processing rather than chunk storage is consistent with the workings of a probabilistic grammar. Given this, and the fact that frequency effects have been observed where they were not expected (Section 3.1), we would not subscribe to the view that frequency effects are evidence of or reliable diagnostics of unit storage.

3.3 Which frequency measure is ideal for predicting entrenchment?

A key question that has received attention only recently (Divjak 2008, Wiechmann 2008, Gries 2013, Schmid and Kuchenhoff 2013, Divjak under review, Levshina under review) is which frequency measure or family of measures is best suited to predict entrenchment? Do different frequency measures correlate with different incarnations of entrenchment (as summarized in Section 2.1)? Issues that are currently debated in assessments of the usefulness of existing frequency measures include the uni- or bi-directionality of the measure, and the inclusion of contingency information and the relevance of statistical null-hypothesis testing.

Earlier experimental work supports association measures (Gries et al. 2005, Ellis and Ferreira-Junior (2009), Ellis and Simpson-Vlach (2009), Colleman and Bernolet (2012). However, research contrasting association measures and conditional probabilities (Divjak 2008; under revision; Levshina under review; Wiechmann 2008; Blumenthal-Drame 2012; Shaoul 2012) shows that conditional probabilities are the favored predictors for a range of linguistic behaviors. Wiechmann (2008), for example, surveyed a wide range of association measures and tested their predictivity using data from eye-tracking during sentence comprehension. The best measure at predicting reading behavior was minimum sensitivity. This measure selects the best of the two available conditional probabilities, i.e. $P(\text{verb}|\text{construction})$ and $P(\text{construction}|\text{verb})$.

Recent studies have compared uni-directional probability measures to bi-directional measures; while the former calculate, for example, $P(\text{verb}|\text{construction})$ or how likely the construction is given a verb, the latter would supplement this information with a calculation of how likely a verb is given the construction and compute both $P(\text{verb}|\text{construction})$ and $P(\text{construction}|\text{verb})$. Divjak (2008; under revision) obtained sentence acceptability ratings on dispreferred and often low frequency Polish combinations of verbs and constructions. Levshina (under review) used gap filling and sentence production tasks on the Russian

ditransitive. Both these studies surveyed a number of association measures, including conditional probabilities, and found that uni-directional probability measures explained behavioral performance at least as well as bi-directional measures. In a similar vein, Blumenthal-Drame (2012) studied the processing of complex word forms in English, using a variety of tasks and both reaction time as well as fMRI measurements. Her conclusion was that (log) relative frequencies (the ratio between surface (root + affix) and base (root) frequencies) predict entrenchment best. Moreover, none of the probability-based measures that outperformed the others on the tasks described above related observed to expected frequencies in order to perform null-hypothesis statistical significance testing. The information gained from relating observed to expected frequencies the way this is done in statistics may have low psychological relevance to speakers.

3.4 The importance of context

Seminal studies from the 1970s, such as Biederman et al (1973), demonstrated that objects are recognized faster and more accurately when accompanied by contextual information. This translates straightforwardly to language, and linguists have indeed focused on frequency effects in language varying in size from phonological to morphological and syntactic contexts. Even disciplines that have been preoccupied with frequency counts, such as corpus linguistics, have borne this principle in mind. Indeed, core concepts in corpus linguistics are collocations, i.e. words that are regularly used together giving rise to an association, and colligations, where a lexical item is linked to a grammatical one. It therefore comes as a surprise to linguists that psychologists interested in language have long focused on words in isolation. Yet behavioral evidence is accumulating that supports linguists' intuitions. One example comes from ERP studies of word processing in sentence context. The magnitude of the N400 component (meaning a negative voltage occurring 400 ms after presentation of a word) indicates difficulty integrating a word with its sentence context. Very large N400s occur for words that are anomalous in their sentence context. N400 wave forms are influenced by word frequency, being largest for very low frequency words. This suggests that contextual integration is most difficult for rare words. However, this frequency effect is strongest at the beginning of a sentence and diminishes for successive words in a semantically congruent sentence (but not a scrambled sentence; van Petten 1993). In van Petten's (1993) study, by the 5th word of a sentence, the N400 frequency effect had disappeared. This suggests that when sufficient context has been encountered, low frequency words are no more difficult to integrate into their context than are high frequency words.

3.5 Is frequency the most important factor for creating entrenched representations?

Following work in the cognitive sciences, we suggest that the answer to this question be 'no.' Frequency is an important contributor, but the relevance of a stimulus for learners' goals may be more important than frequency per se. Entrenchment can occur without repetition frequency, since robust memories can be formed with single-trial learning. A language example is fast mapping, whereby children and adults infer the meaning of a word from context (Carey and Bartlett 1978). But a strong mental representation will be formed from a single instance only in special cases, such as those associated with intense emotions.

Future work on frequency that draws on insights from research on learning, memory and attention and contrasts frequency with salience will no doubt shed light on this question.

4. References

- Adelman, James S., Gordon D. A. Brown, and Jose F. Quesada 2006 Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science* 17:814-823.
- Alegre, M. and P. Gordon. 1999 Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40: 41-61.
- Ambridge, Ben 2013 How do children restrict their linguistic generalizations: an ungrammaticality judgment study. *Cognition* 125:49-63.
- Ambridge, Ben and E. Lieven 2011 *Child language acquisition. Contrasting theoretical approaches*. Cambridge: Cambridge University Press.
- Anderson, John R., and Lael J. Schooler 1991 Reflections of the environment in memory. *Psychological Science* 2: 396 - 408.
- Arnon, Inbal and Neal Snider 2010 More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62:67-82.
- Baayen, R. Harald 2010 Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon* 5: 436-461.
- Baayen, R.H., T. Dijkstra and R. Schreuder 1997 Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language* 37: 94-117.
- Baayen, R. Harold and Peter Hendrix 2011 Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. LSA workshop 'Empirically examining parsimony and redundancy in usage-based models', January 2011.
- Baayen, R. H., Wurm, H. L. and Aycok, J. 2007 Lexical dynamics for low-frequency complex words. A regression study across tasks and modalities. *The Mental Lexicon* 2(3): 419-463
- Baayen, R.H. and Michael Ramscar 2015 xxx
- Balota, David A., Maura Pilotti and Michael J. Cortese 2001 Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition* 29: 639-647.
- Balota, David A., Melvin J. Yap, Michael J. Cortese, and Keith A. Hutchison, Michael J. Cortese, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, and Rebecca Treiman 2007 The English lexicon project. *Behavior Research Methods* 39: 445-459.
- Bannard, Colin, and Danielle Matthews 2008 Stored word sequences in language learning the effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19(3): 241-248.
- Bar, Moshe (ed.) 2011 *Predictions in the brain. Using our past to generate a future*. Oxford: OUP.
- Barlow, H. B. 1990 Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research* 30: 1561-1571.
- Biederman, Irving, A.L. Glass, and E.W. Stacy 1973 Searching for objects in real-world scenes. *Journal of Experimental Psychology* 97: 22-7.
- Blumentahl-Drame, Alice 2012 *Entrenchment in usage-based theories. What Corpus Data Do and Do Not Reveal About The Mind*. Berlin: De Gruyter.

- Bornkessel-Schlesewsky, Ina, and Matthias Schlewsky 2013 Reconciling Time, Space and Function: A New Dorsal-ventral Stream Model of Sentence Comprehension. *Brain and Language* 125(1): 60–76
- Boyd Jeremy K. and Adele E. Goldberg 2011 Learning what not to say: the role of statistical preemption and categorization in “a”-adjective production. *Language* 81 (1): 1-29.
- Braine, M. D. S. and Brooks, P. J. 1995 Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello and W. E. Merriman (eds.), *Beyond names for things: young children's acquisition of verbs*, 353-376. Hillsdale, NJ: Erlbaum
- Brysbaert, Marc, and Boris New 2009 Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41:977-990.
- Bybee, Joan 1995 Regular morphology and the lexicon. *Language and Cognitive Processes* 10: 425-455.
- Bybee, Joan 2007 *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- Bybee, Joan and Paul Hopper (eds) 2001 *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.
- Bybee, Joan, and Thompson, S. 2000 Three frequency effects in syntax. *Berkeley Linguistics Society* 23: 378–388.
- Caldwell–Harris, Catherine L., Jonathan Berant, and Edelman Shimon 2012 Entrenchment of Phrases with Perceptual Identification, Familiarity Ratings, and Corpus Frequency Statistics. In: D. Divjak and S. T. Gries (eds), *Frequency effects in language representation. Volume 2*, 165-194. Berlin: Mouton de Gruyter.
- Carey, Susan and Elsa Bartlett 1978 Acquiring a single new word. *Papers and Reports on Child Language Development* 15: 17-29.
- Carpenter, Shana K., Nicholas J. Cepeda, Dough Rohrer, Sean H.K. Kang, and Harold Pashler 2012 Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Researcher* 24: 369-378.
- Cattell, James M. 1886 The time it takes to see and name objects. *Mind* 41: 63-65.
- Chomsky, Noam 1965 *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Church, Kenneth W. and Patrick Hanks 1990 Word association norms, mutual information and lexicography. *Computational Linguistics* 16: 22-29.
- Cohen, Neal J., and Larry R. Squire 1980 Preserved learning and retention of pattern analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science* 210: 207-210.
- Colleman, Timothy and Bernolet, Sarah 2012 Alternation biases in corpora vs. picture description experiments: DO-biased and PD-biased verbs in the Dutch dative alternation. In: D. Divjak and S. T. Gries (eds), *Frequency effects in language representation Volume 2*, 87-125. Berlin: Mouton de Gruyter.
- Croft, William and Cruse, Alan 2004 *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Cutler, Anne 1981 Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition* 10: 65-70.
- Daelemans, Walter and Antal Van den Bosch. 2005 *Memory-based language processing*. Cambridge: Cambridge University Press.

- Davies, Mark. 2010 Corpus of Contemporary American English. Available from <http://www.americancorpus.org/>
- De Smet, H., and H. Cuyckens 2007 Diachronic aspects of complementation: Constructions, entrenchment and the matching-problem. In: Christopher M. Cain and Geoffrey Russom (eds), *Studies in the History of the English Language III: Managing Chaos: Strategies for Identifying Change in English*, 1–37. Berlin: Mouton de Gruyter.
- Diessel, Holger 2007 Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25: 108–127.
- Divjak, Dagmar (under revision) Pattern transparency and conditional probabilities affect acceptability of verbs in low frequent syntactic alternants.
- Divjak, Dagmar 2008 On (in)frequency and (un)acceptability. In Barbara Lewandowska-Tomaszczyk (ed.), *Corpus Linguistics, Computer Tools and Applications - state of the art*, 1-21. Frankfurt: Peter Lang.
- Dunning, Ted 1993 Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics* 19 (1): 61-74.
- Ellis, Nick C 2002 Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24:143–188.
- Ellis, N. C. and Ferreira-Junior, F. 2009 Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics* 7: 188-221.
- Ellis, N. C. and Simpson-Vlach, R. 2009 Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory* 5: 61-78.
- Elman, J. L. 1993 Learning and development in neural networks: The importance of starting small. *Cognition* 48(1): 71-99.
- Erker D. & G. R. Guy 2012 The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3): 526-557.
- Evert, Stefan 2005 The Statistics of Word Co-occurrences: Word Pairs and Collocations. Ph.D. Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart. Retrieved from <http://www.stefan-evert.de/PUB/Evert2004phd.pdf> on June 28, 2013.
- Fernandez Monsalve, I., S.L. Frank and G. Vigliocco 2012 Lexical surprisal as a general predictor of reading time. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 398-408. Avignon, France: Association for Computational Linguistics.
- Francis, W.N., and H. Kucera. 1982 *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton-Mifflin.
- Gardner, Michael K., E. Z. Rothkopf, Richard Lapan, and Toby Lafferty 1987 The word frequency effect in lexical decision: Finding a frequency-based component. *Memory and Cognition* 151:24-28.
- Grainger, Jonathan 1990 Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language* 29: 228–244.
- Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13: 403-437.
- Gries, Stefan Th. 2010. Dispersions and adjusted frequencies in corpora: further explorations. In Stefan Th. Gries, Stefanie Wulff, and Mark Davies (eds.), *Corpus linguistic applications: current studies, new directions*, 197-212. Amsterdam: Rodopi.

- Gries, Stefan Th., Beate Hampe, & Doris Schönefeld 2005 Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4): 635-676.
- Grill-Spector, Kalanit, Richard Henson, and Alex Martin 2006 Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences* 10: 14-23.
- Gupta, Prahlad 2012 Word Learning as the Confluence of Memory Mechanisms: Computational and Neural Evidence. In Miriam Faust (Ed.), *Handbook of the Neuropsychology of Language, Volume 1*. Oxford: Wiley.
- Gurevich, Olya, Matt Johnson and Adele E. Goldberg 2010 Incidental verbatim memory for language. *Language and Cognition* 2 (1): 45-78.
- Handl, Sandra. 2011. *The Conventionality of Figurative Language: A Usage-based Study*. Tübingen: Narr.
- MacKay, Donald G. 1982 The problems of flexibility, fluency, and speed accuracy trade-off in skilled behavior. *Psychological Review* 89: 483-506.
- Hale, J. 2001 A probabilistic early parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, 1–8. Stroudsburg, PA.
- Hauk, Olaf, Matthew H. Davis, M. Ford, Friedemann Pulvermüller, and William D. Marslen-Wilson 2006 The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage* 30: 1383-1400.
- Hebb, Donald O. 1949 *The organization of behavior*. New York: Wiley and Sons.
- Hills, Thomas T., Josita Maouene, Brian Riordan, and Linda B. Smith 2010 The associative structure of language and contextual diversity in early language acquisition. *Journal of Memory and Language* 63: 259-273.
- Hills, Thomas T., Michael N. Jones, and Peter M. Todd 2012 Optimal foraging in semantic memory. *Psychological Review* 119: 431-440.
- Howes, D. H. and R.L. Solomon 1951 Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology* 416: 401-410.
- Johns, Brendan T., Thomas M. Gruenenfelder, David B. Pisoni and Michael N. Jones 2012 Effects of word frequency, contextual diversity, and semantic distinctiveness on spoken word recognition. *Journal of the Acoustical Society of America* 132: EL74-EL80.
- Jones, Michael N., Johns, Brendan T., and Gabriel Recchia 2012 The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology* 66: 115–124.
- Jurafsky, Dan 1996 A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20: 137–194.
- Langacker, Ronald W. 1987 *Foundations of Cognitive Grammar, Volume I, Theoretical Prerequisites*. Stanford, California: Stanford University Press.
- Lee, Chia-Lin and Federmeier, Kara D. 2012 In a word: ERPs reveal important lexical variables for visual word processing. In Miriam Faust (Ed.), *The Handbook of the Neuropsychology of Language, Volume 1*, 184-208. Oxford: Wiley-Blackwell.
- Levshina, Natalia (under review) Associations between constructions and collexemes revisited: convergent evidence of divergent knowledge.
- Levy, R 2008 Expectation-based syntactic comprehension. *Cognition* 106: 1126–1177.
- Lieven, E. 2010 Input and first language acquisition: Evaluating the role of frequency. *Lingua* 120: 2546-2556.
- Lyne, A. A. 1985 Dispersion. In: *The vocabulary of French business correspondence*, 101–124. Geneva, Paris: Slatkine-Champion.

- Hasher, Lynn and Rose T. Zacks. 1984 Automatic processing of fundamental information: the case of frequency of occurrence. *American Psychologist* 39: 1372.
- MacWhinney, Brian 1998 Models of the emergence of language. *Annual Review of Psychology* 49: 199-227.
- Matthews, Danielle and Grzegorz Krajewski 2015 First language acquisition. xxx
- McClelland, James L. and David E. Rumelhart 1981 An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 88:375-407.
- McDonald, Scott A. and Richard Shillcock 2001 Rethinking the word frequency effect. The neglected role of distributional information in lexical processing. *Language and Speech* 44(3): 295-323.
- Monsell, Stephen 1991 The nature and locus of word frequency effects in reading. In D. Besner and G.W. Humphreys Ed., *Basic processes in reading: Visual word recognition*, 148-197. Hillsdale, NJ: Lawrence Erlbaum.
- Pecina, Pavel 2009 *Lexical Association Measures: Collocation Extraction*. Studies in Computational and Theoretical Linguistics, volume 4. Praha: UFAL.
- Perea, Manuel, Ana Paula Soares, and Montserrat Comesaña 2013 Contextual diversity is a main determinant of word-identification times in young readers. *Journal of Experimental Child Psychology* 116(1): 37-44.
- Pollack, Irwin, Herbert Rubenstein, and Louis Decker 1959 Intelligibility of known and unknown message sets. *The Journal of the Acoustical Society of America* 31: 273-279.
- Pullum, Geoffrey K., and Barbara C. Scholz 2002 Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 18:9-50.
- Rayner, Keith and Susan A. Duffy 1986 Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition* 14: 191-201.
- Rogers, Timothy T. and James L. McClelland 2004 *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT
- Saffran, Jenny R. 2003 Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science* 12: 110-114.
- Saffran, Jenny R., Richard N. Aslin, Elissa L. Newport 1996 Statistical learning by 8-month-old infants. *Science* 274: 1926-1928.
- Schmid, Hans-Jörg 2000 *English abstract nouns as conceptual shells. From corpus to cognition*. Berlin - New York: Mouton de Gruyter.
- Schmid, Hans-Jörg 2010 Does frequency in text really instantiate entrenchment in the cognitive system? In: Dylan Glynn and Kerstin Fischer, eds., *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 101-133. Berlin: Walter de Gruyter.
- Shaoul, Cyrus and Chris Westbury 2011 Formulaic sequences: Do they exist and do they matter? *The Mental Lexicon* 6(1): 171-196.
- Solan, Zach, David Horn, Eytan Ruppín, and Shimon Edelman 2005 Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences* 102: 11629-11634.
- Stefanowitsch, Anatol, and Stefan Th. Gries 2003 Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 82: 209-243.

- Tremblay Antoine, Bruce Derwing, Gary Libben, Chris Westbury 2011 Processing advantages of lexical bundles: evidence from self-paced reading and sentence recall tasks. *Language Learning* 61: 569–613.
- Van Petten, Cyma 1993 A comparison of lexical and sentence-level context effects and their temporal parameters. *Language and Cognitive Processes* 8: 485-532.
- Whaley, C.P. 1978 Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior* 17: 143-154.
- Wiechmann, Daniel 2008 On the Computation of Collostruction Strength. *Corpus Linguistics and Linguistic Theory* 42: 253-290.

Dagmar Divjak and Catherine L. Caldwell-Harris
The University of Sheffield (Sheffield, UK) and Boston University (Boston MA, USA)