Andreas Baumann, Klaus Hofmann*, Anna Marakasova,
Julia Neidhardt and Tanja Wissik

Semantic micro-dynamics as a reflex of occurrence frequency: a semantic networks approach

Abstract: This article correlates fine-grained semantic variability and change with measures of occurrence frequency to investigate whether a word's degree of semantic change is sensitive to how often it is used. We show that this sensitivity can be detected within a short time span (i.e., 20 years), basing our analysis on a large corpus of German allowing for a high temporal resolution (i.e., per month). We measure semantic variability and change with the help of local semantic networks, combining elements of deep learning methodology and graph theory. Our micro-scale analysis complements previous macro-scale studies from the field of natural language processing, corroborating the finding that high token frequency has a negative effect on the degree of semantic change in a lexical item. We relate this relationship to the role of exemplars for establishing form–function pairings between words and their habitual usage contexts.

1 Introduction

Semantic change is among the most conspicuous forms of diachronic variation found in language. When Romeo kills Tybald in the wake of a "nice […] quarrel" (*Romeo and Juliet*, Act 3, Scene 1), the average modern reader will invariably pause and

*Corresponding author: Klaus Hofmann, Department of English and American Studies, University of Vienna, Vienna, Austria, E-mail: klaus.hofmann@univie.ac.at
Andreas Baumann, Department of European and Comparative Literature and Language Studies (EVSL), University of Vienna, Vienna, Austria
Anna Marakasova and Julia Neidhardt, Vienna University of Technology (TU Wien), Vienna, Austria
Tanja Wissik, Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH), Austrian Academy of Sciences, Vienna, Austria

contemplate what the word *nice* might have meant in Shakespeare's times (namely, 'foolish, whimsical, trivial'), as the use of the word in the play does not match its modern denotation. Similarly, when in contemporary parlance every item or action worth commending in the slightest degree is termed *awesome*, this does not evoke the image of a ravaging natural disaster like "a storm of awesome destructive power" (COCA) reported in 1963. Linguists look back on a long tradition of classifying semantic changes like these according to scope (narrowing, widening), evaluation (amelioration, pejoration), or in terms of mechanisms of conceptual transfer (metaphor, metonymy) (Blank and Koch 1999; Bloomfield 1933; Paul 1995 [1880]; Ullmann 1962). However, due to its complexity and the fickle nature of semantics in general, many of the factors involved in semantic change are still not very well understood. In particular, *predicting* change has remained elusive.

This study seeks to contribute to our understanding of the parameters conditioning and influencing diachronic variation of meaning by investigating the link between semantic change and occurrence frequency. Inspired by the achievements of recent large-scale studies employing state-of-the-art methods from the field of natural language processing (Cafagna et al. 2019; Haider and Eger 2019; Hamilton et al. 2016; Rodina et al. 2019; Tahmasebi 2018; Turney and Mohammad 2019), we approach word semantics from a distributional perspective. That is, we assume that a word's meaning is (at least in part) defined and can be expressed by the linguistic context it customarily occurs in. We use embedding-based semantic networks to capture these contexts in a very large corpus of German and compare the networks diachronically. We particularly focus on the way that semantic change relates to microscopic variations in usage on a small temporal scale. Our principal research questions are the following: Does the degree of small-scale semantic variability of lexical items depend on their occurrence frequency? Does small-scale variability in usage translate into frequency-dependent semantic change?

The remainder of this article is structured as follows: Section 2 outlines our theoretical approach within a Cognitive Linguistic framework. Section 3 describes our data and the methods we use for constructing semantic networks and analyzing differences between them. Section 4 reports the results of the study, which are then discussed within the framework of Cognitive Linguistic theory in Section 5. Section 6 concludes the article.

2 Theoretical approach

This study subscribes to the research agenda of Cognitive Linguistics. From among the various related and partly overlapping theoretical positions subsumed under

that banner, we regard the following five tenets as the theoretical cornerstones of our research:

(1) Linguistic knowledge is *usage-based.* Experience with linguistic events in natural language usage gives rise to and continuously shapes mental representations of linguistic categories. Categories exist in the form of emergent and intrinsically variable clusters of mental associations between form and meaning/function.

(2) Linguistic representations on different levels (phonology, morpho-syntax, lexicon) are *exemplar-based.* Exemplars are equivalent to the traces of individual linguistic experiences impressed into memory. Categories emerging from exemplars are sensitive to statistical properties of usage, notably frequency of occurrence and contextual diversity. Due to their probabilistic structure, exemplar clouds naturally display *prototype effects.*

(3) According to the *distributional hypothesis*, similarity in lexical meaning can be expressed through co-occurrence patterns between linguistic items in sufficiently large and representative natural language datasets.

(4) *Semantic change* is the result of the inherent communicative flexibility and context-dependency of lexical meaning. Both synchronic polysemy and diachronic change are brought about by mechanisms such as metaphor, metonymy and pragmatic inference in language usage.

(5) *Occurrence frequency* is a key factor in language processing, variation and change.

2.1 Linguistic knowledge is usage-based

A linguistic category is created through the establishment of a mental pairing of (usually, but not necessarily) arbitrary acoustic, gestural or graphemic form with conceptual or grammatical function (Croft 2001; Goldberg 1995, 2006; Hoffmann and Trousdale 2013; Langacker 1987). This is achieved through entrenchment, that is, the mental integration of a form–function mapping as a result of repeated perception of the association between form and function in usage (Bybee 2007). Langacker (1987: 59) characterizes the process as follows:

> 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 […].

Thus, a linguistic category emerges cumulatively as positive evidence for the form–function mapping becomes available to the language learner (Behrens 2009; Bybee

2010; Ellis et al. 2013). Conversely, lack of attestations results in failure to establish a mental link and therefore a lack of acceptability (Ambridge et al. 2015; Braine and Brooks 1995). New, unattested links can nevertheless be formed by reanalysis or creative and context-dependent usage (cf. Section 2.4).

2.2 Linguistic representations are exemplar-based and display prototype effects

Exemplar theory (Bybee 2007; Goldinger 1998; Nosofsky 1988a; Pierrehumbert 2001, 2003) serves as a bridge between the usage-based approach and the theoretical requirement that models of linguistic representation be compatible with domain-general psychological mechanisms. The concept has been particularly prominent in phonology, where categories are taken to represent an individual's aggregated experience with speech sounds. This takes the form of "a large cloud of remembered tokens of that category […] organized in a cognitive map, so that memories of highly similar instances are close to each other and memories of dissimilar instances are far apart" (Pierrehumbert 2001: 140). The idealized 'default value' of a phonological category, then, will be the value at the center of the phonetic region covered by the exemplars in the phonetic memory space. Statistically, this corresponds to the mode of the probability distribution of its occurrence in usage. Due to the focus on the probabilistic grounding of categories in linguistic experience, the main contribution of exemplar theory has been to account for observed frequency effects at the interface between phonology and morpho-syntax (Bybee 2007; Bybee and Thompson 1997) as well as the considerable amount of phonetic detail remembered by language users (Goldinger 1998; Jusczyk 2000; Miller 1997; Nygaard et al. 2000).

Exemplar models naturally accommodate *prototype effects*. This includes the common finding that some items are judged to be better or more basic representatives of a category than others and that these items are also retrieved more efficiently (Rosch 1975, 1978). The prototype structure of categories is a foundational notion in cognitive theory, particularly in semantics (Coleman and Kay 1981; Craig 1986; Geeraerts 1985, 1997). Its central claim is that concepts themselves display a radial or network-like structure, with prototypical members at the center and less proto-typical ones at increasing distances from it. Group membership is established through similarity, which is judged based on a variable selection of 'family resemblance' features rather than a clearly delineated set of 'necessary and sufficient' ones (Rosch and Mervis 1975). Polysemy relations are also accommodated by this model: Basic meanings of a word form the prototype center, from which links extend to additional meanings at the margins (Cuyckens 1995; Lakoff 1987).

In analogy to the exemplar model in phonology, lexical concepts can also be thought of as exemplar clouds made up of rich, multi-modal memory traces of

situations, objects, actions, or, broadly speaking, contexts in which a speaker has experienced the lexeme being applied. Exemplar theory would suggest that the placement of a category member within the graded prototype structure is to a large extent determined by frequency of occurrence (Nosofsky 1988b). A type of bird that is commonly encountered in a speaker's native land (e.g., titmouse in Central Europe), may be more prototypical to a speaker than other species (e.g., parrots), as they may hear the term relatively more often applied to the former than the latter. Several factors may qualify the relationship between frequency and typicality, however. Prototypes do not need to match actual exemplars but may often take the shape of abstractions combining salient features from various exemplars (Divjak and Arppe 2013). Also, it has been reported that typicality correlates with age of acquisition, so that typical exemplars tend to be ones that are acquired earlier (Holmes and Ellis 2006), even allowing for the fact that age of acquisition and frequency are closely correlated (Kuperman et al. 2012).

If one accepts the programmatic non-modularity of linguistic knowledge in Cognitive Linguistics (Langacker 1987), the structure of lexical representations should in principle be no different from phonological or grammatical categories. While the variability captured by the cloud-like structure of phonological exemplars is mostly restricted to the formal side (i.e., the range of encountered articulations), variability in lexical semantics primarily concerns the functional side of the form–meaning mappings as referents denoted by the same item can differ from each other (e.g., *leg* 'limb' vs. 'support for table top' vs. 'stage of journey or process'). Based on the assumptions of the exemplar model, conceptual knowledge can afford to be inclusive, capturing rich multi-modal information about referents and usage contexts. Thus, the exemplar-based representation of a lexical category is one in which "the tokens of words […] are represented in memory along with the situations they have been associated with in experience" (Bybee 2013: 64, cf. Croft 2021: 245–272). This is not to deny that abstraction and schematicity form an integral part of semantic representations. Rather, the existence of frequency and prototype effects suggests that an exemplar-like architecture lies at the basis of such abstractions and has a decisive influence on them.

2.3 Lexical meaning reflects distributional properties of words in language use

Not all lexical items refer to tangible objects that can be perceived with one's senses. The meaning of many items may be almost entirely based on intralinguistic associations without direct recourse to extralinguistic reality. Yet, this does not seem to pose much of a problem to either communication or learning. It is a common

experience that the meaning of a new word can be inferred from its linguistic context. Crucially, this is often possible even when the item relates to a relatively concrete referent. Observations such as these have prompted some linguists and lexicologists to formulate strong assumptions about the role of a lexeme's distribution and collocational preferences in the representation of its meaning, best encapsulated by Firth's famous quotation: "You shall know a word by the company it keeps!" (Firth 1957: 11).

The approach that regards collocational preferences as central to conceptual structure has become known as the *distributional hypothesis*. This line of thinking integrates well with an exemplar-based viewpoint, since the linguistic context of words forms part of "the situations they have been associated with in experience" (Bybee 2013: 64) and is stored in episodic memory alongside it. There are of course legitimate reasons to ask whether linguistic context should be considered the *only* kind of context that is relevant for conceptual representations. There is an inherent circularity in distributional models that try to explain the meaning of words solely by their associations with other words, which is sometimes referred to as the *grounding problem*. Studies informed by embodiment theory have pointed to sensomotoric experience to circumvent this problem (Barsalou 2010; Pulvermüller 2005). Nonetheless, it has been demonstrated that a concept's perceptual features, i.e., the aspects that refer to non-linguistic modes of cognition such as vision, sound, touch and smell, can be successfully predicted based on linguistic input data alone (Johns and Jones 2012; Louwerse and Connell 2011; cf. Arbib et al. 2014). In other words, even though distributional theories may fall short of explaining how lexico-semantic representations form in the mind, they can still be used to model them, including properties that form part of the extralinguistic context (for hybrid approaches in the distributional vs. embodied debate, cf. Louwerse 2007; Vigliocco et al. 2009).

The distributional hypothesis has given rise to a thriving research tradition in computer linguistics, where it serves as the theoretical foundation for a range of methods for measuring and comparing lexical meaning(s) based on large collections of natural language data (Griffiths et al. 2007; Kutuzov et al. 2018; Landauer and Dumais 1997; Mikolov et al. 2013). The present paper takes inspiration from this rich fund of methods (cf. Section 3).

2.4 Semantic change originates in language usage

Since the linguistic categories in the minds of speakers emerge from usage, variation and change are considered integral features of language within the usage-based model (Ellis and Larsen-Freeman 2006; Langacker 1999: 91–146). Semantic change occurs when the conventionalized mapping between a lexical item and its referent(s)

is altered in some way over time. Depending on whether the alteration takes place on the formal or functional end of the pairing, the change may be said to be onomasiological (what is *X* called) or semasiological (what does *X* mean), although changes that purely affect only one dimension while leaving the other untouched are probably rare (e.g., lexical replacement of body part vocabulary). For methodological reasons we are chiefly concerned with the semasiological perspective here (cf. Section 3).

On the one hand, semantic changes happen through rather subtle shifts in the pragmatic aspects of a word's meaning as it is construed through usage. This phenomenon has been analyzed in terms of implicature and inference: Part of the pragmatic context in which an item is used is reinterpreted as an essential part of its conceptual structure (Traugott 2018; Traugott and Dasher 2002). On the other hand, semantic change may also come about as the result of creative language use for the purpose of effective communication (Geeraerts 1997). Thus, established form–meaning pairings can be extended to include new concepts that are related to the original one(s) by similarity (through metaphor) or contiguity (through metonymy), so that some conceptual features are shared between them (Lakoff and Johnson 1980; Sweetser 1990). Language is constantly used creatively in this way, to express concepts that still lack conventional linguistic form (e.g., the novel use of *file* for a section of memory space with a retrievable path on a computer hard drive) or to frame messages in specific ways by highlighting some features at the expense of others (e.g., as part of political rhetoric, Lakoff 2002). Mechanisms such as invited inferencing and metaphorical extension create polysemies that are themselves organized in terms of prototype relations. The most conspicuous semantic changes are those where the prototype center shifts over time so that a former peripheral conceptual cluster becomes more central (Györi 2002), not least due to changes in the relative frequency of usage. An example is the word *bead*, whose Middle English ancestor *bede* meant 'prayer'. The original link between the two is metonymic: Beads assembled on a rosary are used to count prayers in Christian practice.

Assuming, as we do, that conceptual structure is highly inclusive in nature, these points suggest that semantic change is gradient, ranging from subtle contextual differences, over fairly independent uses, to perceived unrelatedness. In addition, change is diachronically gradual, as the prototypes inspired by input-based exemplars wax and wane with usage. In terms of a distributional operationalization of conceptual representation, this means that the similarity between the meanings of an item at different points in time can be regarded as the amount of similarity between their collocation networks in a distributional model. This is also the way that semantic similarity will be measured in this study (cf. Section 3).

2.5 The role of frequency in semantic change

Frequency of occurrence has long been recognized as an influential factor in explaining aspects of language behavior and language change. Frequency effects are commonly observed in psycholinguistic experiments, where higher frequency facilitates lexical access, processing speed and accuracy (Ellis 2002; Howes and Solomon 1951; Whaley 1978). This is in line with the predictions made by the exemplar model, according to which the most frequent instantiations of a category are also the most entrenched, located at the center of the exemplar cloud, and therefore accessed more easily (Bybee and Hopper 2001; Ellis et al. 2016: 45–60).

The effects of frequency on diachronic change are variable, depending on the kind of phenomenon in question. On the one hand, high-frequency items are more likely to undergo phonological reduction changes than less frequent ones (Bybee 2007). This often affects grammaticalizing items, such as *gonna* (< *going to*), which are routinely found in prosodically weak positions, but the same kind of routinization has also led to reductions in main class words such as *cupboard*. On the other hand, frequency can have the seemingly opposite effect of preserving forms through time: High-frequency items tend to resist paradigm leveling more than low-frequency items. Thus, irregular verb forms such as *swim–swam–swum* or *eat–ate–eaten* are among the most frequent verbs in English (Bybee 2007). The link between frequency and irregularity had previously been pointed out by Greenberg (1966). A similar phenomenon is observed in rates of lexical replacement, where words for frequent concepts display higher diachronic stability (Lieberman et al. 2007; Pagel et al. 2007). Such effects follow directly from the exemplar model, where high-frequency items can resist paradigmatic analogy forces due to their stable entrenchment within a tight exemplar cloud.

In comparison to phonology and morpho-syntax, the relationship between frequency and semantics had long remained relatively underexplored. This is somewhat surprising, considering that speculations about a possible correlation between semantic change and occurrence frequency go as far back as the writings of Paul (1995 [1880]), who speculated that rarer words display a higher tendency to undergo semantic reanalysis because occasion-bound 'misinterpretations' of such items are less likely to be corrected by repeated exposure to the 'correct' form–meaning mapping.[1] A more recent corpus-driven study by Hamilton et al.

---

1 "Wir werden also eine Art des Bedeutungswandels anzuerkennen haben, die darauf beruht, dass der für die ältere Generation usuellen Bedeutung von der jüngeren eine nur partiell damit über-einstimmende untergeschoben wird. Das Gebiet dieser Art des Wandels werden wir aber auf die selteneren und nicht leicht klar zu fixierenden Begriffe einzuschränken haben, da bei anderen die allmähliche Korrektur nach dem bestehenden Usus nicht ausbleiben kann." (Paul 1995 [1880]: 86)

(2016) has investigated the relationship between frequency and semantic change from a distributional perspective, drawing on two of the largest available historical language datasets, (Google N-Grams and COHA). They find that the rate of semantic change – measured as the cosine distance between the semantic vectors representing a word's meaning at different points in time – is indeed negatively correlated with token frequency. This holds true even when a measure of a word's degree of polysemy is included in the regression model. Polysemy itself seems to have the opposite effect of increasing the rate of semantic change, despite frequency and polysemy being positively correlated (i.e., other things being equal, the more frequent a word, the more polysemous it is, cf. also Casas et al. 2019; Jager et al. 2016).

Further word embeddings-based studies of different semantic change measures (Cafagna et al. 2019; Haider and Eger 2019; Rodina et al. 2019; Tahmasebi 2018) generally support the findings in Hamilton et al. (2016). However, a few studies employing short-term time spans (as we do here, cf. Section 3), have indicated that high word frequency does not necessarily imply stability of meaning or context-specific word usage (Del Tredici et al. 2019; Kahmann et al. 2017; Vylomova et al. 2019). Others suggest that the effect of frequency on the rate of semantic change could also be explained as model artifacts (Dubossarsky et al. 2016, 2017). There is a risk to overinterpret differences in word meaning representations that actually stem from noise in the data, although the issue is rather intricate as diachronic frequency movements may themselves contribute to such noise. Another study suggests a new approach for investigating the relationship between semantic change and frequency by comparing semantic divergence in a set of cognates across languages (Uban et al. 2019). That study finds an effect opposite to that observed by Hamilton et al. (2016). However, the authors concede that their word frequency ranks are not contemporaneous with their measures of semantic similarity. In contrast, Pagel et al. (2007) have adopted an onomasiological perspective, concluding that concepts that have high associated word frequencies in modern languages are more likely to be expressed through the cognates inherited from the common proto-language, while less frequent concepts are more prone to lexical replacement.

The present paper is strongly influenced by Hamilton et al. (2016) as we also study the effect of frequency on semantic change. In addition, an explicit aim is to integrate the findings with a Cognitive Linguistic framework. Since we assume, in line with current theory, that semantic change is historically gradual and fundamentally based on small usage variations that play out on the pragmatic level, we hypothesize that the effects found by Hamilton et al. (2016) should also be visible at much smaller time scales (e.g., months and years instead of centuries) given adequate data resolution. To accommodate this difference in perspective, our approach departs from previous research regarding the data used as well as the methods applied for modeling meaning, measuring semantic change over time and statistical analysis.

3 Data and methods

3.1 Corpus data

We base our analysis on the Austrian Media Corpus (AMC, Ransmayr et al. 2017), which is currently the largest diachronic corpus representing the Austrian variety of German. It collects texts (journalistic prose) from the majority of Austrian print and online media and spans a period of more than three decades. In total, it consists of about 11 billion word tokens (of 8.5 billion word-form types) distributed over 45 million texts. For our study, we use all print media texts from 1998 to 2018 (because sub-corpus sizes are considerably smaller for previous years). Within this period, yearly sub-corpus sizes range between 1.3 and 2.1 million texts. All texts are time-stamped and linguistically pre-processed (in particular, part-of-speech-tagged).

The size and granularity of the corpus are ideal for our purposes, since we are interested in semantic-contextual developments within a short time span. To ensure that derived time series are based on a large enough number of data points, a high temporal resolution is necessary. Fortunately, the AMC is large enough to extract reliable word-frequency estimates for monthly sub-periods: There are between 100,000 and 170,000 texts *per month*, so that the sizes of the 252 monthly sub-corpora range from 25 to 43 million word tokens (to provide some context: The well-studied Corpus of Historical American English, COHA, accommodates about 23 million word tokens, on average, *per decade*; Davies 2010). The sub-corpus sizes are certainly sufficient for the computational steps applied in the analysis (cf. Section 3.5).

The corpus represents a relatively confined variety (Austrian German) and genre (journalistic prose). Although this is an incidental characteristic of the corpus, which we have selected for its size and granularity, we do not consider this a disadvantage since any cognitively motivated relationships between frequency and contextual usage should be expected to be independent from the genre they are detected in. Clearly, frequency effects in contextual shifts operate subconsciously regardless of the text type. If such frequency effects can be detected in such a relatively standardized context (written, non-colloquial, journalistic language) they should be expected to extend to less normalized domains of linguistic usage.

3.2 Data set

Addressing the relationship between contextual-semantic change and occurrence frequency in a systematic way requires a data set that is representative of the predictor variables under investigation. Our starting point is a representative

sample of lexical items, for which we then construct semantic network representations. Meaning variation and change will be measured by comparing the network representations associated with these lexical targets (cf. Section 3.3).

Since words in natural languages are distributed across the frequency spectrum in a notoriously uneven way (Zipf 1935, 1949), we may not succeed in detecting the expected effects if we relied on a purely random sample. Additionally, the occurrence frequency of any linguistic item should not be regarded as an invariant feature either, a fact that the ever-increasing number of historical linguistics studies documenting the waxing and waning of morphemes, constructions and lexemes testifies to. The lexical items' actual semantic content, on the other hand, is of relatively minor importance for structuring our sample. It may safely be assumed that a good deal of semantic space will be covered by the sample given that it is large enough.[2] Based on these considerations, we stratified our sample according to three frequency-related measures: (1) mean monthly frequency, (2) diachronic frequency development, and (3) frequency fluctuation. The corpus lexicon was divided into tertiles on each of these measures. For each combination of sampling strata, the maximal number of words allowing an equal number of items across all combinations was randomly selected, summing up to just below 3,000 items altogether (37 lemmas × 3 mean FRQ tertiles × 3 FRQ development tertiles × 3 FRQ fluctuation tertiles × 3 PoS = 2,997; see Supplementary Material A.1 for further details on the sampling procedure).[3]

---

2 As our sample is based on forms and not meanings, we cannot rule out that there will be words with unrelated meanings (i.e., homonyms) in the sample. This bears some potential for skewing the data if homonymy relates systematically to frequency of use. However, excluding homonymic forms from the sample also brings with it certain problems, both conceptual as well as practical. On the one hand, it has been noted that the distinction between polysemy and homonymy is often a gradient rather than a categorical one (e.g., Deane 1988). In our own data, classifying a lexical form as having related or unrelated meanings is certainly not always easy. For example, the form *ablaufen* can mean 'to expire/to go bad' or it can be used in the phrase *jemandem den Rang ablaufen* 'to outperform somebody'. The polysemic links, however, reside in the constituent morphemes *ab-* and *laufen* 'run' rather than the complex form itself, producing meanings that are related only distantly and indirectly (cf. Plag 2003; Plank 2010). On the other hand, *lesen* 'to read' or 'to collect/harvest' is treated as homonymic in dictionaries (e.g., DUDEN, s.v. *lesen*), while its meanings are historically related and it cannot be ruled out that some speakers may still perceive it that way. We have therefore decided to leave potentially homonymic forms in the sample. However, a systematic check has shown that only a small minority of the words in the sample receive homonym-style entries in a dictionary.
3 During the construction of the lexical networks, a handful of items had to be discarded as the distributional data was not sufficient for network construction, so the actual number of analyzed items is slightly lower.

3.3 Lexical networks

We make use of network graphs to represent the context-based semantics of a target word and to measure semantic change. Graph-based methods have been employed in various areas of natural language processing (NLP). Lexical networks in particular have been used for unsupervised word sense induction (Akkasi and Snajder 2021; Biemann 2006; Hope and Keller 2013; Navigli and Lapata 2009), word sense disambiguation (Bevilacqua and Roberto 2020; Dorow et al. 2004), lexical semantic relatedness computation (Hughes and Ramage 2007), automatic synsets induction (Ustalov et al. 2017), unsupervised POS-tagging (Degórski 2013), unsupervised lexical acquisition (Biemann 2006; Widdows and Dorow 2002) or induction of lexical taxonomies in a semi-automated mode (Navigli et al. 2011). Recently, network-based algorithms have been applied to efficiently infer various similarity measures by computing node embeddings (Kutuzov et al. 2019). The structure of lexical networks also makes it possible to cluster nodes with high interconnectedness into sense-level groups, which helps determine the type of change (i.e., narrowing, broadening, birth, death), in addition to detecting the fact or degree of change (Jana et al. 2019; Mitra et al. 2014, 2015; Tahmasebi and Risse 2017). That is, lexical networks implicitly contain information about the semantic structure of words (and, in particular, polysemy). A promising application of graph theory in semantics can be found in a series of works on word usage graphs (McCarthy et al. 2016; Schlechtweg et al. 2021a, 2021b), where human-annotated similarity ratings of corpus examples of target words are used to create evaluation datasets for semantic change research.

    Like many of the studies listed above, we apply an ego network approach: Each individual network embodies the contextual semantics of a target word based on its distributional characteristics in a given monthly sub-corpus. The nodes of the networks represent lexemes that have similar distributional usage characteristics to the target, while the edges in turn represent similarity relations among them (Figure 1). Thereby, we derive a densely interconnected and thus computationally stable representation of a target words' semantic content based on the corpus data (cf. Supplementary Material A.2 for further details on network construction). Meaning change can then be detected by comparing ego-networks of the same target word across different time periods. The general idea of this approach is straightforward: Target words whose ego-networks are structured similarly (i.e., have similar nodes and edge structure) are also semantically similar.[4]

---

4 Again, our approach views semantic change from an expressly semasiological perspective. One could also imagine a study design where networks are generated from corpus data which are not defined by their attachment to a specific target (i.e. networks that are not ego networks) but some other criterion such as network size or internal coherence. These could then be clustered and
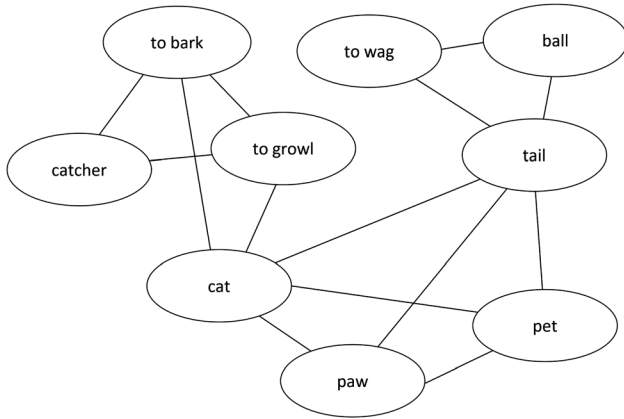
Figure 1: Hypothetical ego network for the target item *dog*. Nodes represent lexemes with similar distributional usage characteristics, edges represent similarity relations among them, which may be paradigmatic (e.g., *to bark* vs. *to growl*) or syntagmatic (e.g., *to bark* vs. *cat*). Since all nodes are by definition connected to the target, the target itself is excluded from the network.

3.4 Network distance as a measure of contextual variability and change

Change in contextual usage is captured by the quantified difference between two networks, each representing the semantics of the target word at a given point in time. The information about which nodes in a network are connected by an edge can be expressed in the form of a data matrix (Figure 2). We then determine how dissimilar two networks are by using the Frobenius norm, which defines a distance metric quantifying the difference between two network matrices in Euclidean space (Horn and Johnson 2012).

There are two ways in which change in contextual usage can be measured (Figure 3): (a) by measuring the distance between semantic representations of consecutive periods (t_i vs. t_{i+1}); or (b) by measuring the distance between each of the individual representations in the time series from a common point of origin

---

compared diachronically to find out whether these networks associate more or less stably with specific lexical forms over time. Such an onomasiological approach, which would be grounded in function rather than form, is however much more computationally intensive as it is less targeted. Generally speaking, starting at the functional end of the form–function pairing makes data collection more difficult, but see Croft (2021: 245–272) for ideas on how to study exemplar semantics from an onomasiological perspective by controlling the situations that experiment participants are asked to verbalize.
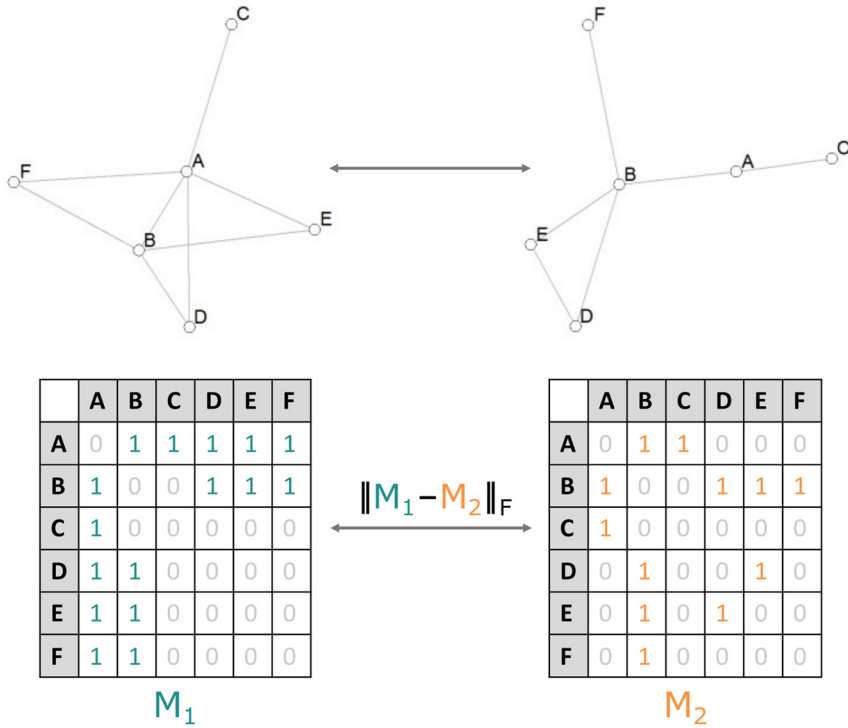
Figure 2: Distance between two adjacency matrices M1 and M2 of two networks in terms of the Frobenius norm. For a matrix M, the Frobenius norm is defined as $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$ (effectively the Euclidean norm of the flattened matrix).

(t_0 vs. t_i). The former has been employed by Hamilton et al. (2016). Strictly speaking, therefore, the results of that study imply that occurrence frequency is negatively correlated with semantic *variability* rather than semantic *change* in the traditional sense, as each of their data points represents one decade-long time interval and not a time series-style progression of change from a fixed starting point. In the present study, both perspectives are employed to address slightly different aspects of the relationship between contextual-semantic change and frequency of occurrence.

The item *transparent* serves to illustrate the amount of change that can be expected from our data (Figure 4). Like its English counterpart, the adjective has two metaphorically related meanings in German, viz. 'clear, see-through' (as in *transparent glass*) and 'obvious, understandable, open to public scrutiny' (as in *transparent hiring practices*). The network from the beginning of the period distinguishes clearly between those two meanings. One usage cluster relates to the context of architectural or possibly textile
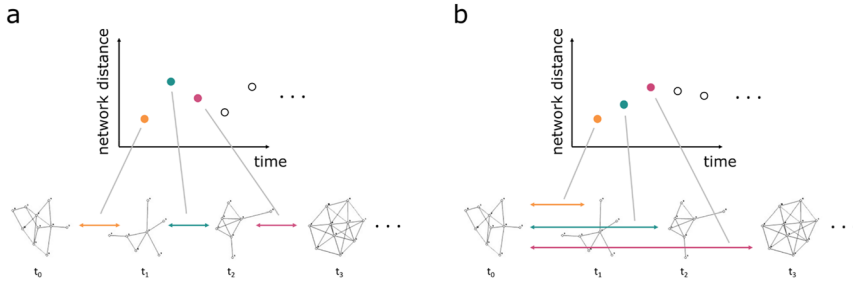
Figure 3: Two ways of measuring change in contextual usage: (a) Distances between pairs of networks in consecutive years measure the degree of contextual variability. (b) Pairwise distances between any network and the network in the first period measure the extent to which a word shifts away from its original usage at t_0.

design (*Farbe* 'color', *optisch* 'visual(ly)', *hell* 'bright', *elegant* 'elegant'), while the other, although somewhat vague, seems to capture the context of public or managerial processes (*Anforderung* 'requirement', *effizient* 'efficient', *Struktur* 'structure').

In contrast, the network from the end of the period distinguishes between the context of digital technology and consumerism (*Anwendung* 'application', *Verbesserung* 'improvement', *Konsument* 'consumer', *innovativ* 'innovative' etc.)
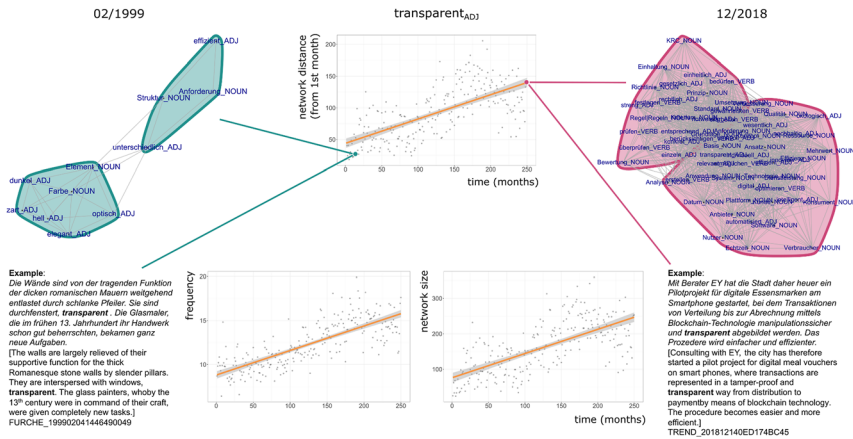

Figure 4: Development of the German adjective *transparent* ('transparent'). Mid-upper panel: change in contextual usage (measured as distance from common origin, i.e., network in the first month) depending on time measured in months. The increasing line indicates that *transparent* shifts away from its original usage. Left (petrol) and right (magenta) panel: example networks and contexts for an early (02/1999) and a late period (12/2018), respectively. Mid-lower panel: the adjective *transparent* has increased in frequency (left) and network size (right).

and the public/legal/regulatory sphere (*gesetzlich* 'legal', *Regel* 'rule, regulation', *prüfen* 'probe, inspect', *Richtlinie* 'guideline', etc.). Crucially, in 2018 the literal meaning seems to have become marginalized by the densely represented metaphorical use, to the extent that it is not recognized by the underlying embedding algorithm anymore. This may well reflect an ongoing semantic shift away from the literal to the metaphorical meaning in Austrian German usage.[5]

That the two example networks are representative of a larger trend for the item *transparent* can be seen in the upper middle panel. Here, change in usage is measured in terms of how distant each network in the time series is from the very first network of the investigation period, as described under (b) above. Thus, each data point stands for the semantic distance between the network representing a given month and the network at t_0, i.e., January 1998. It emerges that network distance as a measure of change in contextual usage increases over time.

Following the literature and assuming that the observed subtle alterations in usage represent the incipient stages of more drastic semantic change, we would expect this to coincide with low frequency of occurrence. However, the relationship is complicated by the fact that the target word's frequency does not remain constant, but is itself undergoing a marked upward shift. In other words, changes in contextual usage and frequency coincide in this case.

The same is true of network size, i.e., the number of nodes populating the network, as is reflected in the noticeably larger network from 2018. Although it would be a gross simplification to equate network size with the range of meanings linked to an item, it may nevertheless be regarded as a proxy for usage flexibility, in the sense that a higher number of nodes suggests a higher number of items which the target is related to through usage. To some degree, network size may also be determined by the embedding algorithm itself, so it is worth treating this feature with caution. Regardless of its interpretation, network size needs to be controlled for to get a clear picture of the effects of frequency on contextual change. In the case of *transparent*, all three variables seem to be positively correlated. This would clearly conflict with earlier findings should it turn out to be a general trend throughout our data set.

3.5 Description of variables and modeling procedure

In order to disentangle the relationships among these variables, we proceed as follows: First, we relate the MEAN CONTEXTUAL VARIABILITY of our target items, measured as the

---

5 It is worth noting that the literal meaning has a native competitor in *durchsichtig*, differing in register but not in denotation, while there is no clear single candidate for the more abstract metaphorical meaning. This may be a factor conducive to semantic change in this item.

network distance between consecutive months averaged across the investigation period (cf. Figure 3a above), to MEAN FREQUENCY, on the one hand, and MEAN NETWORK SIZE, on the other (both averaged across the period). PART OF SPEECH is also taken into account as a predictor. All three continuous variables are heavily skewed towards the lower end of their respective ranges. For MEAN FREQUENCY this was to be expected, keeping in mind the uneven Zipfian distribution of lexical items across the frequency spectrum (Zipf 1935), despite our efforts to stratify the sample with respect to frequency (cf. Section 3.2). Similar to utterance frequency, corpus size and network size are positively correlated (cf. Supplementary Material A.3 for a detailed discussion of corpus size and frequency effects on network similarity).[6] Hence, we normalized all network-based measures, i.e., network size and distances between networks, with respect to corpus size (per month, employing a multiplicative constant of 10^7 to avoid numerical issues). To improve the model's quality, MEAN CONTEXTUAL VARIABILITY, MEAN NETWORK SIZE and MEAN FREQUENCY were transformed logarithmically. The two continuous predictors MEAN NETWORK SIZE and MEAN FREQUENCY are positively correlated. In order to avoid problems of multicollinearity, this needs to be monitored closely for the regression model.

In a second step, we consider change in usage not as variability between consecutive period intervals, but as movement away from the semantic representation at point t_0 (cf. Figure 3b above). We do this by calculating for every month the distance of the targets' networks from the very first month of the investigation period, resulting in a time series of distances from the first month (again normalized with respect to corpus size). Then, we fit a generalized additive model (Wood 2006) to each time series, in order to obtain a smoothly curved distance trajectory. After that, we compute the area under this curve and define this area as CONTEXTUAL CHANGE. The rationale behind this operationalization is this: A word with contextual semantics moving away from the original meaning will show a strongly increasing distance trajectory (e.g., *transparent* in Figure 5). In contrast, a word with relatively stable contextual semantics throughout the observation period like the adjective *fraglich* ('questionable') is expected to show a relatively flat trajectory not far from the first month (i.e., close to a distance of zero). However, words could also first move away from their original contextual meaning, but, after some time, return again. This is illustrated by the word *fett* ('fat, adipose, bold') in Figure 5 (besides 'adipose', the adjective *fett* can also mean 'drunk', 'great', and can be used similarly to the affirmative expression *awesome* in English; apparently, some of these uses were more prominent one decade ago than they are now). Obviously, *fett* has changed to a larger extent than *fraglich*, although both end up close to their initial contextualized meaning. Simply computing the distance between the first and the final month would

---

6 We would like to thank one of the reviewers for addressing the need to consider corpus size when analyzing lexical networks.
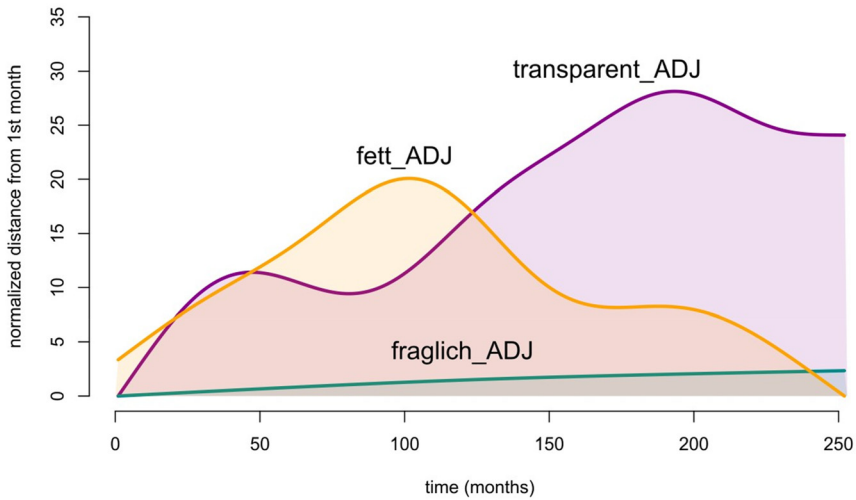
Figure 5: Trajectories of contextual deviation from the first month for three different words: *transparent* ('transparent'), *fett* ('fat'), and *fraglich* ('questionable'). Depending on the semantic dynamics, trajectories show differential areas below them.

hide dynamics like this, as would the computation of the slope (or some other measure of average increase) of the trajectory. The area under the trajectory, however, does capture such subtle differences between the respective dynamics.[7]

In addition, we also account for the changeable character of frequency as well as network size by computing Spearman correlations for their diachronic trajectories, defining this as FREQUENCY CHANGE and NETWORK SIZE CHANGE. In short, we model the extent to which changes in contextual usage can be explained by changes in occurrence frequency while controlling for changes in network size. The variable PART OF SPEECH is kept unchanged.

We opt for linear regression as the method of statistical analysis and compute two models. In both models, each data point represents one word. In the first model, MEAN CONTEXTUAL VARIABILITY depends on MEAN FREQUENCY, MEAN NETWORK SIZE and PART OF SPEECH. In addition, since we were mostly interested in frequency effects on semantic dynamics, we implemented multiplicative interactions between MEAN FREQUENCY and MEAN NETWORK SIZE, as well as between MEAN FREQUENCY and PART OF SPEECH as control terms. We also weighted MEAN CONTEXTUAL VARIABILITY by the reciprocal of its margin of error (based on the 95 % confidence intervals of the estimated means) in order to account for the fact that some trajectories of pairwise differences fluctuate to a larger

---

7 The relevance of these differences when studying semantic dynamics was rightfully stressed by one of the reviewers.

extent than others (which immediately affects the accuracy of the mean estimates). All predictors are implemented as fixed effects.[8]

The second model features CONTEXTUAL CHANGE as outcome variable depending on the predictors (fixed effects) FREQUENCY CHANGE, NETWORK SIZE CHANGE and PART OF SPEECH. In addition, we include interactions between FREQUENCY CHANGE and NETWORK SIZE CHANGE, as well as between FREQUENCY CHANGE and PART OF SPEECH as control terms (Table 1).

To discriminate among the range of models that are possible given our predictor sets and interaction terms, we employ the *Akaike information criterion* (AIC) (Burnham et al. 2011; Johnson and Omland 2004). AIC is preferred over plain goodness-of-fit measures (such as R^2) as it penalizes models for their complexity, thus striking a balance between complexity and descriptiveness. Additionally, we test for multi-collinearity using the *variance inflation factor* (VIF), a measure quantifying the extent to which a regression term is affected by multicollinearity (Grueber et al. 2011). All statistical operations are carried out with *R* (R Core Team 2023, version 4.2.2).

4 Results

4.1 Mean contextual variability

The first set of results concerns the extent to which MEAN CONTEXTUAL VARIABILITY of the target words is explained by MEAN FREQUENCY, MEAN NETWORK SIZE and PART OF SPEECH.[9] The best-performing model (as determined by AIC and VIF analyses) is summarized in Table 2 and Figure 6.

The model coefficients confirm our expectations. When network size is controlled for, contextual variability is inversely dependent on frequency, i.e., items

---

8 Note that by having summary measures, i.e., aggregated values like MEAN FREQUENCY, for every single word, we avoid the implementation of random effects as in Hamilton et al. (2016). We opted for this procedure for two reasons: First, we consider aggregated scores to be easier to interpret; second, a random effect 'word type', as implemented in Hamilton et al. (2016), would fail to grasp the temporal structure of the underlying time series (i.e., the model would be invariant with respect to temporal permutations). As one reviewer has rightly pointed out, random effects preserving the temporal structure could in principle be implemented as random smooths in generalized additive models (where each word type represents one level of the random variable). However, this could lead to over-specification (Baayen and Linke 2021). Most importantly, though, although being a useful alternative for studying effects on semantic changes from one period to the next (i.e., CONTEXTUAL VARIABILITY), a random-effects architecture would make it impossible to study *long-term* change as operationalized by CONTEXTUAL CHANGE in our second model. This is so since the random effect (be it a random intercept, slope, or smooth) would collect exactly those data points from which CONTEXTUAL CHANGE would be derived.
9 Distance comparisons involving networks of zero size were excluded. A handful of items were excluded altogether due to scarcity of viable networks.

Table 1: Breakdown of all variables considered in this study together with their respective derivation and interpretation. Dependent variables are highlighted in bold. All numeric variables were standardized (*z*-transformed) before entering the regression models.

Variable	Derivation for each word	Interpretation
MEAN FREQUENCY	Arithmetic mean (across all months) of normalized token frequency (subsequently log transformed)	Average usage frequency in the observation period
MEAN NETWORK SIZE	Arithmetic mean (across all months) normalized number of nodes (subsequently log transformed)	Average number of contextually associated words in the observation period
MEAN CONTEXTUAL VARIABILITY	Arithmetic mean (across all months) of normalized consecutive network distances (via Frobenius norm) as in Figure 3a (subsequently log transformed)	Average monthly fluctuation in contextual usage in the observation period
FREQUENCY CHANGE	Spearman correlation coefficient of time and normalized token frequency	Growth/decline in usage frequency
NETWORK SIZE CHANGE	Spearman correlation coefficient of time and normalized network size (number of nodes)	Growth/decline in the number of contextually associated words
PART OF SPEECH	PoS tag extracted from corpus (noun, verb, adjective; adjective as default class)	Word class
CONTEXTUAL CHANGE	Area under the trajectory of normalized network distance from origin (network in first month)	Rate of contextual shift away from original usage at t_0 over the observation period

that are more frequent tend to show less contextual variability between months than less frequent items. Put differently, the more frequent the item, the more stable its usage context. Although frequency is significant as a main effect, its effect size is very small. It has to be kept in mind that both the predictor and the response variables have been transformed in preprocessing, so the actual relationship between the two is not linear. The coefficients and effect sizes need to be viewed in this light.

In contrast, network size has a strong positive effect on contextual variability and is by far the major contributor to the general predictiveness of the model. As a main effect, this relationship is rather difficult to interpret, due to the fact that network size itself factors into the computation of both the networks and the network distances, which is why it needed to be added as a control variable to begin with.[10] Importantly, network size cannot be regarded as a straightforward measure of a target's semantic

---

10 The Frobenius norm requires two compared matrices to be of equal size. Since we did not constrain the size of the networks in order to meet this expectation, as this would have leveled down the networks to size 0, we instead supplied the smaller of two compared networks with dummy edges of weight 0. This very likely inflated the main effect of network size on network distance.

Table 2: Coefficient estimates of the first model (MEAN CONTEXTUAL VARIABILITY depending on MEAN FREQUENCY, MEAN NETWORK SIZE and PART OF SPEECH) together with their 95 % confidence intervals. All statistically non-trivial effects are highlighted in bold. All coefficients are rounded for the second decimal place.

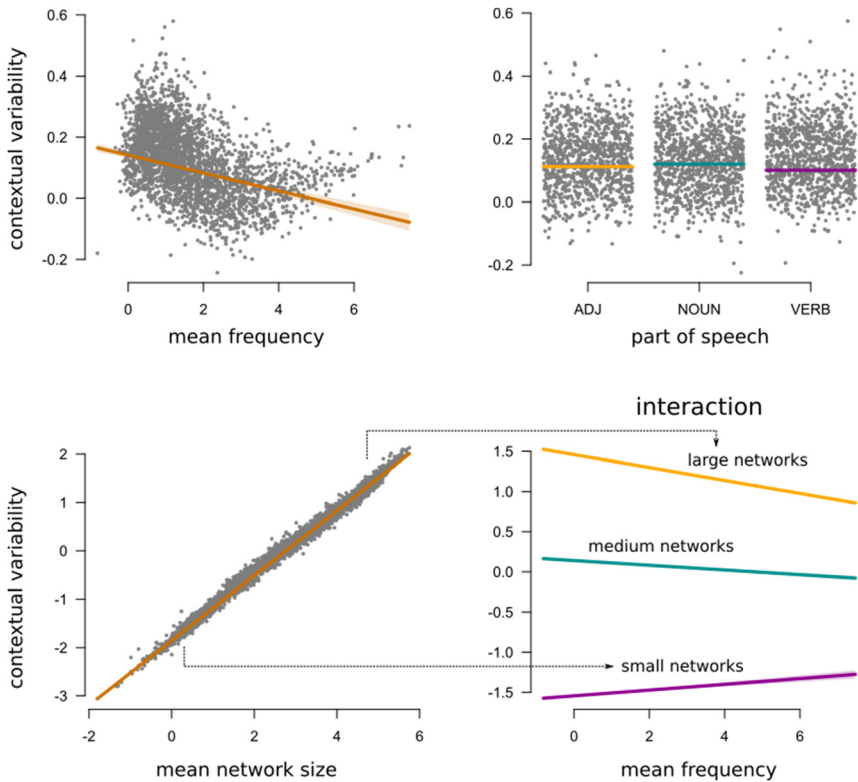| Predictor/interaction	Estimate	95 % CI
(Intercept)	0.01	(0.01, 0.01)
MEAN FREQUENCY	**−0.03**	**(−0.04, −0.03)**
MEAN NETWORK SIZE	**1.04**	**(1.04, 1.05)**
PART OF SPEECH: NOUN	0.01	(−0.00, 0.02)
PART OF SPEECH: VERB	**−0.01**	**(−0.02, −0.00)**
MEAN FREQUENCY × MEAN NETWORK SIZE	**−0.05**	**(−0.06, −0.05)**

Figure 6: Effects on MEAN CONTEXTUAL VARIABILITY. Each data point represents one target word. Upper panel: MEAN FREQUENCY has a negative impact on MEAN CONTEXTUAL VARIABILITY (left). Verbs show lower MEAN CONTEXTUAL VARIABILITY (right). Lower panel: MEAN NETWORK SIZE has a positive effect on MEAN CONTEXTUAL VARIABILITY (left). Interaction of MEAN NETWORK SIZE and MEAN FREQUENCY (right). Medium and large networks display a negative effect of frequency on contextual variability. In small networks, frequency correlates positively with contextual variability.

scope. At the same time, it cannot be ruled out that semantic scope may still be reflected by network size to some extent. In this regard, inspecting the interaction between network size and frequency, rather than the main effect, may be more informative, since it allows us to partition the range of network sizes into smaller slices and observe the effect of frequency in each slice individually. This suggests that the negative effect of frequency is not constant across words with different average network sizes, but that it is stronger in words with larger networks.

Three tentative conclusions can be drawn from this part of the analysis: (1) in line with our expectations, words that occur more often have more stable usage contexts; (2) verbs display slightly lower contextual variability than nouns and adjectives; (3) the effect of frequency is somehow moderated by the size of the networks.

4.2 Contextual change

Up to this point, the measures for contextual variability, frequency and network size have been treated as averages across the whole period. This does not only entail a considerable amount of fuzziness in the data, but leaves the crucial question of larger diachronic change, in contrast to monthwise variability, unaddressed. In the second part, therefore, we consider CONTEXTUAL CHANGE as a function of FREQUENCY CHANGE while controlling for NETWORK SIZE CHANGE and PART OF SPEECH. We do this to do justice to the fact that frequency and network size are themselves changeable (cf. Section 3.5).

The best-performing model (as determined by AIC and VIF analyses) is summarized in Table 3 and Figure 7.

With regard to the main effects, the results from this analysis are similar to what we have seen in the previous section. FREQUENCY CHANGE is negatively related to CONTEXTUAL CHANGE. In other words, items that have become less frequent over the investigation period tend to display more change in their usage contexts than items that become more frequent. Once again, the effect size is small, but the effect is statistically significant. NETWORK SIZE CHANGE is positively related to CONTEXTUAL CHANGE. PART OF SPEECH also emerges as a significant factor in this model. In particular, nouns seem to have slightly higher rates of change than verbs (with adjectives placed in-between).

More intriguing than the individual factors' main effects are the interactions among the variables, in particular the interaction between FREQUENCY CHANGE and NETWORK SIZE CHANGE. As can be seen in Figure 7, the negative effect of FREQUENCY CHANGE on CONTEXTUAL CHANGE is particularly marked in items whose networks have become

Table 3: Coefficient estimates of the second model (CONTEXTUAL CHANGE depending on FREQUENCY CHANGE, PART OF SPEECH, NETWORK SIZE CHANGE and all interactions) together with their 95 % confidence intervals. Significantly non-trivial effects are highlighted (bold).

Predictor/interaction	Estimate	95 % CI
(Intercept)	−0.15	(−0.21, −0.09)
FREQUENCY CHANGE	**−0.07**	**(−0.12, −0.01)**
NETWORK SIZE CHANGE	**0.21**	**(0.17, 0.25)**
PART OF SPEECH: NOUN	**0.18**	**(0.10, 0.26)**
PART OF SPEECH: VERB	**−0.11**	**(−0.19, −0.02)**
FREQUENCY CHANGE × NETWORK SIZE CHANGE	**0.22**	**(0.19, 0.25)**
FREQUENCY CHANGE × PART OF SPEECH: NOUN	−0.02	(−0.02, 0.06)
FREQUENCY CHANGE × PART OF SPEECH: VERB	0.03	(−0.05, 0.11)

smaller over time (i.e., words with contracting networks). In such networks, an increase in frequency leads to more stable usage contexts, while decreasing frequency entails changing usage contexts. These effects match the expectations derived from the literature. Interestingly, however, the effect is reversed in items whose networks become larger over time. That is, in expanding networks, growth in frequency leads to larger amounts of contextual change.

The intricate relationship between frequency and network size in this model points to the possibility that network size might encode more than a computational confound after all. If it was nothing else than an artifact introduced by the method, we would not expect it to moderate the effect of frequency in the way that we have observed. On the contrary, the fact that change in frequency profiles seems to have different effects in expanding or contracting networks suggests a more ambivalent role of occurrence frequency when it comes to stabilizing the semantics of a word.

5 Discussion

The results of the quantitative analysis are in line with our expectations. High occurrence frequency of lexical items has a stabilizing effect on their semantic representations. This is true both on the level of the variability of usage contexts between consecutive months and on the level of longitudinal change in usage contexts over several years. In conjunction, the two measures can be regarded as capturing a micro-perspective on contextual-semantic change, complementing
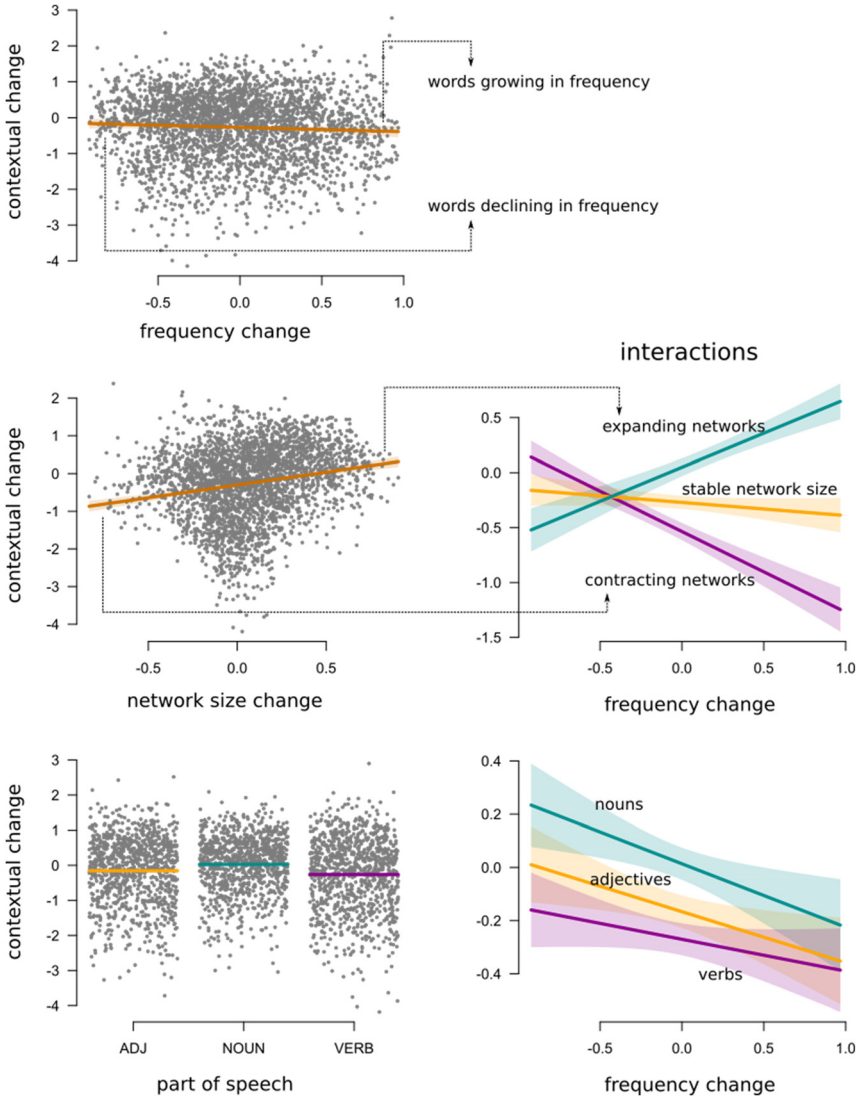
Figure 7: Effects on CONTEXTUAL CHANGE. Top panel: FREQUENCY CHANGE (rate of growth) has a negative impact on CONTEXTUAL CHANGE. Middle panel: NETWORK SIZE CHANGE (rate of growth) and CONTEXTUAL CHANGE are positively correlated. The effect of FREQUENCY CHANGE on CONTEXTUAL CHANGE is affected considerably by PART OF SPEECH and reversed for expanding networks. Bottom panel: effect of PART OF SPEECH on contextual change: nouns display slightly higher rates of contextual change than adjectives and verbs show lower contextual change than adjectives. The negative effect of FREQUENCY CHANGE on CONTEXTUAL CHANGE is not affected by PART OF SPEECH.

existing studies based on data spanning several centuries (Cassani et al. 2021; Dubossarsky et al. 2017; Hamilton et al. 2016).

The role of occurrence frequency for stabilizing semantic meaning has been speculated about at least since the 19th century, when Hermann Paul sought to explain the correlation by referring to quasi-normative pressures exerted by exposure to the 'usual meaning' ("der […] usuellen Bedeutung") of frequent items and the relative lack of such pressures on rarer items. The metaphorical frame Paul uses in his colorful description conjures up the image of subversive youths surreptitiously supplanting ("unterschieben") the word semantics of their elders with partly shifted ones, which frequent lexical items are more resistant to because they are less variable in their 'existing usage' ("bestehenden Usus") (Paul 1995 [1880]).

Leaving aside his insinuations regarding the subversive inclinations of the younger generation, Paul's description is remarkably in line with current linguistic theory, and the results of the present study, in particular. It touches on the usage-based character of semantic representations as well as the link between frequency and semantic variability. The mechanism by which frequency exerts its stabilizing effect, on the other hand, remains vague in his account.

An exemplar-based conceptualization of linguistic knowledge can be helpful to fill this gap. According to the exemplar model, frequency is related to the degree of cognitive entrenchment: the more frequent words are in the ambient language, the more stably they become entrenched in the minds of speakers. With regard to semantic representations, this is conceived of as the accumulation of exemplars. Lexical exemplar clouds consist of, "the tokens of words […] represented in memory along with the situations they have been associated with in experience" (Bybee 2013: 64). Each encounter of a word in context contributes to the construction of its mental representation, either by reinforcement or by slight reconfiguration of the memory traces that encode conceptual knowledge from linguistic and sensomotoric experience. The context in which words are uttered thus becomes the crucial ingredient for the construction of semantic knowledge, in all its linguistic and multi-modal richness.

Analogous to the implementation of contextual usage as networks in this study, linguistic knowledge is itself often conceived of as an internal network connecting various levels of traditional linguistic analysis (Diessel 2019). In fact, a network-like architecture underlying linguistic and other cognitive abilities is suggested by what is known about the functioning of neurons in the brain. This has been the inspiration for connectionist models of cognition (Rumelhart et al. 1986; Westermann et al. 2009), also known as 'neural networks', which have since become one of the cornerstones of

deep learning algorithms used in natural language processing and elsewhere.[11] From this perspective, the strengthening of exemplar clouds can be recast as the strengthening of the connections in the network(s) of linguistic representation, which is in turn linked to conceptual knowledge about the world.

The tighter these cognitive associations are, the more likely it is that a given communicative situation will trigger the use of the word most strongly associated with it (say, the use of *transparent* for describing the see-through character of glass within an architectural context, rather than, for example, the synonym *durchsichtig*). Vice versa, hearing a word uttered will invoke the conceptual knowledge that is most strongly associated with it through usage. It is not hard to imagine how prototype effects develop out of the frequency-sensitive strengths of such associations. These associations are obviously subject to change, but the stronger the connections between words and their conceptual content, the smaller the amount of variability in usage, and the slower the rate of semantic change. To fully demonstrate the strength of association, one would ideally have to relate semantic variation not only to the occurrence frequency of the target, but also to the frequencies of any potential lexical competitors for the same function, i.e., words that (partly) share the same referent. This is so because entrenchment is an intricate phenomenon which applies to the form–function pairing in a number of distinguishable ways (cf. Geeraerts 1997). Among other things, it must be assumed that entrenchment is strongest when the frequency of the target is high while at the same time the frequencies (and overall number) of the target's lexical competitors are low – in the (unrealistic) extreme, the relationship would be bi-unique. The example of *transparent* and *durchsichtig*, is a case in point. Determining the relevant set of lexical competitors for each of our target items extends well beyond the confines of our paper, but this onomasiological aspect is certainly something that can and should be developed further in future research.[12]

Apart from the methodological aspects, one of the main contributions of this paper is that the effect of frequency on semantic stability, previously observed for long time periods in various languages, can also be detected on a much smaller time scale, within only 20 years, and even in the subtle variations in contextual usage between consecutive months. This hardly involves sweeping semantic changes at the level of English *gay* or *awful* (Hamilton et al. 2016). Instead, our investigation has

---

11 It needs to be emphasized that the nodes in the hidden layers of connectionist models cannot be equated with units of analysis in conventional linguistics. Generally, the hidden layers in neural networks capture generalizations that lie much deeper than what usually counts as units of analysis in linguistics (phoneme, morpheme, word etc.), hence this mode of analysis is also called 'subsymbolic' (cf. Smolensky 1987).

12 We thank one of the reviewers for pointing this out.

focused on slow and subtle contextual oscillations in the words' usages, most likely unnoticed by language users, but which can accumulate and alter semantic meaning in the long run. The fact that the correlation can be detected at all on this microscopic scale can be interpreted as evidence for the existence of a continuum between pragmatics and semantics, between synchronic context-induced variability and diachronic semantic shift. On a broader theoretical level, it further corroborates the view that the formation of linguistic categories is fundamentally based in linguistic usage. The exact mechanisms by which subtle alterations in contextual usage filter into the semantic representations of lexical items are varied. Many of these questions have been addressed in a wide array of mostly qualitative studies in terms of grammaticalization, subjectification and invited inferences, but more research employing state-of-the-art computational methodology is required to investigate these links on a larger scale.

Quantitative explorations into the motivations for semantic change are still relatively rare. In Cassani et al. (2021), words that are learned earlier in language acquisition are found to be less prone to change over generations (cf. Monaghan and Roberts 2019). This would seem to attribute change mostly to the process of first language acquisition, an interpretation that is not dissimilar from Paul's characterisation. In our case, however, it makes relatively little sense to explain the negative correlation between frequency and semantic-contextual change as being mediated by first language learners, considering that our data does not span across consecutive generations of language users. Instead, the fact that similar tendencies can be observed in our data suggests that contextual usage itself is impacted by frequency. In other words, language users tend to use frequent items in less variable ways. Frequent words are more firmly entrenched in the minds of language users, which has a constraining effect on their usage variability. This in turn disincentivizes change.

Apart from these general points, it is worth commenting on some of the more intricate patterns that our investigation has revealed. On the one hand, the analysis of contextual change suggests that nouns are slightly, but significantly, more prone to change than adjectives and verbs, particularly in words with decreasing frequency. These results stand in contrast with previous studies (Dubossarsky et al. 2016) which indicate that nouns are more stable diachronically. There are also good reasons why this should be the case: Nouns are typically acquired earlier than verbs, age-of-acquisition generally being a reliable predictor for diachronic stability (Gentner and Boroditsky 2001; Monaghan 2014). Also, verbs are generally more polysemous than nouns (Gentner and France 1988), and polysemy has been shown to correlate with likelihood of change (Hamilton et al. 2016). So, the finding that nouns actually seem to change *more* than verbs and adjectives is somewhat puzzling. One possible

interpretation would take into account that nouns typically refer to entities that can be seen, touched, smelt etc. In other words, their meanings rely more on the embodied mode of semantic representation (Barsalou 2010; Pulvermüller 2005), which is captured only indirectly in our study, while verbs (and to a lesser extent, adjectives) rely more on distributional linguistic information rather than senso-motoric information. This may go some way towards explaining the larger degree of semantic fluctuation detected for nouns in our data: Because nominal meaning does not depend on the linguistic context to the same extent, the linguistic context can afford to vary more. Such an interpretation would point towards a critical limitation of the distributional semantics approach, calling for further investigation with methods that also monitor embodiment effects, for example by controlling for concreteness (Brysbaert et al. 2014; cf. also Dunn 2015; Reijnierse et al. 2019; Winter and Srinivasan 2022 for a related discussion in conceptual metaphor theory) or perceptual strength (Lynott et al. 2020; cf. also Bruni et al. 2014).

Finally, the finding that the effect of frequency change on semantic-contextual change is moderated by changes in network size is one aspect that particularly stood out. In essence, this suggests that words undergoing a frequency increase resist change mostly in stable or contracting networks. As discussed earlier, network size by itself is at best an indirect indicator of a word's semantic range. However, its moderating effect does imply some connection between the range of linguistic items in a word's habitual usage context and the stabilizing effect of frequency on a word's semantics. The overall effect seems to be one of specialization, i.e., a combination of frequency increase and network contraction coupled with semantic stabilization. The diachronic frequency trajectories associated with semantic narrowing are still awaiting systematic quantitative investigation. However, a possible parallel to our results may be found in construction grammar, more specifically in the different effects of token and type frequency. Type frequency is generally taken as a measure of the productivity of a construction, but it is also related to its schematicity: The larger the number of items eligible for filling a slot, the broader the range of meanings that can be accommodated by it and the wider the range of contexts where it can be used (Barðdal 2008; Bybee and Thompson 1997; De Smet 2020). The flipside of this are strongly entrenched combinatorial patterns, such as idioms and prefabs, which are highly constrained in composition and usage, but may nevertheless be very frequent (Bybee 2006; Erman and Warren 2000; Sinclair 1991). It is possible that the interaction patterns observed in our study reflect a similar dynamic on the level of lexical semantics, where the stabilizing effect of frequency may be limited to words that are rather constrained in their distributional behavior. Once again, rather than providing a conclusive answer, this finding may serve to generate more, and more specific, hypotheses that merit investigation in future research.

6 Conclusion

The aim of this article was to investigate the relationship between meaning and frequency of occurrence. It has done so by comparing NLP-inspired semantic networks representing the contextual usage of a representative sample of target words over a short period of time with high temporal resolution. The results corroborate previous studies that have diagnosed a negative effect of occurrence frequency on the degree of semantic change, that is, words with higher frequencies tend to change less than words that occur less often. We argue that this effect follows rather naturally from a usage-based conception of linguistic competence. As linguistic knowledge is based on experience with language, high frequency of experience with a linguistic sign translates into stronger entrenchment of the form–function pairing in the mental lexicon, which in turn disincentivizes change. The network approach has proved a promising avenue for grasping the complexity of the form–function link in the case of word semantics. In addition, we have demonstrated that variability in usage displays the same frequency-dependency as patterns of diachronic change, thus providing additional evidence for the view that language is constantly reshaped through usage.

Our study faces obvious limitations in terms of the breadth of the data used. We hope that future research can replicate our results on similarly structured data with a high temporal resolution from other languages, ideally incorporating more diverse genres and discourse types. We also recommend that future studies take into account the relationship between frequency and diversity of occurrence contexts (cf. Adelman et al. 2006) in predicting semantic change to probe more deeply into the cognitive mechanisms mediating between linguistic experience and linguistic knowledge. Finally, the onomasiological perspective needs further elaboration. Taking lexical competition for semantic function into account in a more explicit way will improve our understanding of the way in which entrenchment ultimately depends on usage.

Data availability statement

The datasets generated during and/or analysed in the current study are available in the PHAIDRA repository at https://phaidra.univie.ac.at/o:1742666. Additional information about the sampling procedure, network construction, and the relationship between utterance frequency, corpus size and network size can be found in the Supplementary Material published online along with this article.

References

Adelman, James S., Gordon D. A. Brown & José F. Quesada. 2006. Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science* 17(9). 814–823.

Akkasi, Abbas & Jan Snajder. 2021. Word sense induction using leader-follower clustering of automatically generated lexical substitutes. *Expert Systems with Applications* 181. 115162.

Ambridge, Ben, Amy Bidgood, Katherine E. Twomey, Julian M. Pine, Caroline F. Rowland & Daniel Freudenthal. 2015. Preemption versus entrenchment: Towards a construction-general solution to the problem of the retreat from verb argument structure overgeneralization. *PLoS One* 10(4). e0123723.

Arbib, Michael A., Brad Gasser & Victor Barrès. 2014. Language is handy but is it embodied? *Neuropsychologia* 55. 57–70.

Baayen, R. Harald & Maja Linke. 2021. Generalized additive mixed models. In Magali Paquot & Stefan Th. Gries (eds.), *A practical handbook of corpus linguistics*, 563–591. Cham: Springer.

Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam; Philadelphia: John Benjamins.

Barsalou, Lawrence W. 2010. Grounded cognition: Past, present, and future. *Topics in Cognitive Science* 2(4). 716–724.

Behrens, Heike. 2009. Usage-based and emergentist approaches to language acquisition. *Linguistics* 47(2). 383–411.

Bevilacqua, Michele & Roberto Navigli. 2020. Breaking through the 80 % glass ceiling: Raising the state of the art in word sense disambiguation by incorporating knowledge graph information. In *Proceedings of the 58th annual meeting of the Association for Computational Linguistics*, 2854–2864. Association for Computational Linguistics. Available at: https://aclanthology.org/2020.acl-main.255/.

Biemann, Chris. 2006. Chinese Whispers: An efficient graph clustering algorithm and its application to natural language processing problems. In *Proceedings of TextGraphs: The first workshop on graph based methods for natural language processing*, 73–80. Association for Computational Linguistics. Available at: https://aclanthology.org/W06-3812/.

Blank, Andreas & Peter Koch. 1999. Introduction: Historical semantics and cognition. In Andreas Blank & Peter Koch (eds.), *Historical semantics and cognition*, 1–16. Berlin; New York: Mouton de Gruyter.

Bloomfield, Leonard. 1933. *Language*. New York: Allen & Unwin.

Braine, Martin D. S. & Patricia J. Brooks. 1995. Verb argument structure and the problem of avoiding an overgeneral grammar. In Michael Tomasello & William E. Merriman (eds.), *Beyond names for things: Young children's acquisition of verbs*, 353–376. Hillsdale: Erlbaum.

Bruni, Eli, Nam Khanh Tran & Marco Baroni. 2014. Multimodal distributional semantics. *Journal of Artificial Intelligence Research* 49. 1–47.

Brysbaert, Marc, Amy Beth Warriner & Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46. 904–911.

Bullinaria, John A. & Joseph P. Levy. 2007. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods* 39(3). 510–526.

Burnham, Kenneth P., David R. Anderson & Kathryn P. Huyvaert. 2011. AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65. 23–35.

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

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford; New York: Oxford University Press.

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

Bybee, Joan L. 2013. Usage-based theory and exemplar representations of constructions. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of construction grammar*, 49–69. Oxford: Oxford University Press.

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

Bybee, Joan, L. & Sandra Thompson. 1997. Three frequency effects in syntax. *Annual Meeting of the Berkeley Linguistics Society* 23(1). 378–388.

Cafagna, Michele, Lorenzo De Mattei & Malvina Nissim. 2019. Embeddings shifts as proxies for different word use in Italian newspapers. In *Proceedings of the sixth Italian conference on computational linguistics*. Achen: CEUR-WS.org. Available at: https://ceur-ws.org/Vol-2481/paper12.pdf.

Casas, Bernardino, Antoni Hernández-Fernández, Neus Català, Ramon Ferrer-i-Cancho & Jaume Baixeries. 2019. Polysemy and brevity versus frequency in language. *Computer Speech and Language* 58. 19–50.

Cassani, Giovanni, Federico Bianchi & Marco Marelli. 2021. Words with consistent diachronic usage patterns are learned earlier: A computational analysis using temporally aligned word embeddings. *Cognitive Science* 45(4). e12963.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences*. New York: Erlbaum.

Coleman, Linda & Paul Kay. 1981. Prototype semantics: The English word *Lie*. *Language* 57(1). 26–44.

Craig, Colette G. (ed.). 1986. *Noun classes and categorization*. Amsterdam; Philadelphia: John Benjamins.

Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, William. 2021. *Ten lectures on construction grammar and typology*. Leiden; Boston: Brill.

Cuyckens, Hubert. 1995. Family resemblance in the Dutch spatial prepositions *door* and *langs*. *Cognitive Linguistics* 6(2–3). 183–208.

Davies, Mark. 2010. *The corpus of historical American English: COHA*. Brigham: Brigham Young University.

Deane, Paul D. 1988. Polysemy and cognition. *Lingua* 75(4). 325–361.

Degórski, Łukasz. 2013. Fine-tuning Chinese Whispers algorithm for a Slavonic language POS tagging task and its evaluation. In Zygmunt Vetulani (ed.). *Proceedings of the 6th language and technology conference*, 439–443. Available at: http://nlp.ipipan.waw.pl/Bib/dego:13:ltc.pdf.

Del Tredici, Marco, Raquel Fernández & Gemma Boleda. 2019. Short-term meaning shift: A distributional exploration. *Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies*, 1, 2069–2075. Association for Computational Linguistics.

De Smet, Hendrik. 2020. What predicts productivity? Theory meets individuals. *Cognitive Linguistics* 31(2). 251–278.

Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.

Divjak, Dagmar & Antti Arppe. 2013. Extracting prototypes from exemplars: What can corpus data tell us about concept representation? *Cognitive Linguistics* 24(2). 221–274.

Dorow, Beate, Dominic Widdows, Katarina Ling, Jean-Pierre Eckmann, Danilo Sergi & Elisha Moses. 2004. Using curvature and Markov clustering in graphs for lexical acquisition and word sense discrimination. https://doi.org/10.48550/arXiv.cond-mat/0403693.

Dubossarsky, Haim, Daphna Weinshall & Eitan Grossman. 2016. Verbs change more than nouns: A bottom-up computational approach to semantic change. *Lingue e Linguaggio* 25(1). 5–25.

Dubossarsky, Haim, Daphna Weinshall & Eitan Grossman. 2017. Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, 1136–1145. Association for Computational Linguistics.

Dunn, Jonathan. 2015. Modeling abstractness and metaphoricity. *Metaphor and Symbol* 30(4). 259–289.

Eckart, Carl & Gale Young. 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1(3). 211–218.

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(2). 143–188.

Ellis, Nick C. & Diane Larsen-Freeman. 2006. Language emergence: Implications for Applied Linguistics. Introduction to the Special Issue. *Applied Linguistics* 27(4). 558–589.

Ellis, Nick C., Matthew Brook O'Donnell & Ute Römer. 2013. Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning* 63. 25–51.

Ellis, Nick C., Ute Römer & Matthew Brook O'Donnell. 2016. Constructions and usage-based approaches to language acquisition. *Language Learning* 66(S1). 23–44.

Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20(1). 29–62.

Firth, John R. 1957. A synopsis of linguistic theory 1930–1955. In John R. Firth (ed.). *Studies in linguistic analysis*, 1–32. Oxford: Philological Society.

Geeraerts, Dirk. 1985. *Paradigm and paradox: Explorations into a paradigmatic theory of meaning and its epistemological background*. Leuven: Leuven University Press.

Geeraerts, Dirk. 1997. *Diachronic prototype semantics: A contribution to historical lexicology*. Oxford: Clarendon Press.

Gentner, Dedre & Ilene M. France. 1988. The verb mutability effect: Studies of the combinatorial semantics of nouns and verbs. In Steven L. Small, Garrison W. Cottrell & Michael K. Tanenhaus (eds.), *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence*, 343–382. San Mateo, CA: Kaufmann.

Gentner, Dedre & Lera Boroditsky. 2001. Individuation, relativity, and early word learning. In Melissa Bowerman & Stephen Levinson (eds.), *Language acquisition and conceptual development*, 215–256. Cambridge: Cambridge University Press.

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

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

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2). 251–279.

Green, Clarence. 2017. Usage-based linguistics and the magic number four. *Cognitive Linguistics* 28(2). 209–237.

Greenberg, Joseph H. 1966. *Language universals. With special reference to feature hierarchies*. The Hague: Mouton & Co.

Griffiths, Thomas L., Mark Steyvers & Joshua B. Tenenbaum. 2007. Topics in semantic representation. *Psychological Review* 114(2). 211–244.

Grueber, Catherine E., S. Nakagawa, R. J. Laws & Ian G. Jamieson. 2011. Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology* 24. 699–711.

Györi, Gábor. 2002. Semantic change and cognition. *Cognitive Linguistics* 13(2). 123–166.

Haider, Thomas & Steffen Eger. 2019. Semantic change and emerging tropes in a large corpus of new high German poetry. In *Proceedings of the 1st international workshop on computational approaches to historical language change*, 216–222. Association for Computational Linguistics.

Hamilton, William L., Jure Leskovec & Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *Proceedings of the 54th annual meeting of the Association for Computational Linguistics, Volume 1: Long papers*, 1489–1501. Association for Computational Linguistics.

Hoffmann, Thomas & Graeme Trousdale (eds.), 2013. *The Oxford handbook of construction grammar*. Oxford; New York: Oxford University Press.

Holmes, Selina J. & Andrew W. Ellis. 2006. Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition* 13(7–8). 884–910.

Hope, David & Bill Keller. 2013. MaxMax: A graph-based soft clustering algorithm applied to word sense induction. In Alexander Gelbukh (ed.). *Computational linguistics and intelligent text processing. CICLing 2013*. Berlin, Heidelberg: Springer.

Horn, Roger A. & Charles R. Johnson. 2012. *Matrix analysis*. Cambridge: Cambridge University Press.

Howes, Davis H. & Richard L. Solomon. 1951. Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology* 41(6). 401–410.

Hughes, Thad & Daniel Ramage. 2007. Lexical semantic relatedness with random graph walks. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning* (EMNLP-CoNLL), 581–589. Association for Computational Linguistics. Available at: https://aclanthology.org/D07-1061/.

Jager, Bernadet, Matthew J. Green & Alexandra A. Cleland. 2016. Polysemy in the mental lexicon: Relatedness and frequency affect representational overlap. *Language, Cognition and Neuroscience* 31(3). 425–429. https://doi.org/10.1080/23273798.2015.1105986.

Jana, Abhik, Animesh Mukherjee & Pawan Goyal. 2019. Detecting reliable novel word senses: A network-centric approach. *Proceedings of the 34th ACM/SIGAPP symposium on applied computing*, 976–983. New York: Association for Computing Machinery. https://doi.org/10.1145/3297280.

Johns, Brendan T. & Michael N. Jones. 2012. Perceptual inference through global lexical similarity. *Topics in Cognitive Science* 4(1). 103–120.

Johnson, Jerald & Kristian Omland. 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution* 19(2). 101–108.

Jusczyk, Peter W. 2000. *The discovery of spoken language*. Cambridge, Mass: MIT Press.

Kahmann, Christian, Andreas Niekler & Gerhard Heyer. 2017. Detecting and assessing contextual change in diachronic text documents using context volatility. *Proceedings of the 9th international joint conference on knowledge discovery, knowledge engineering and knowledge management (KDIR 2017)*, 135–143. Setúbal: SciTePress. https://doi.org/10.5220/006574001350143.

Kuperman, Victor, Hans Stadthagen-Gonzalez & Marc Brysbaert. 2012. Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods* 44. 978–990.

Kutuzov, Andrey, Lilja Øvrelid, Terrence Szymanski & Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: A survey. In *Proceedings of the 27th international conference on computational linguistics*, 1384–1397. Association for Computational Linguistics.

Kutuzov, Andrey, Mohammad Dorgham, Oleksiy Oliynyk, Chris Biemann & Alexander Panchenko. 2019. Making fast graph-based algorithms with graph metric embeddings. *Proceedings of the 57th annual meeting of the Association for Computational Linguistics*, 3349–3355. Association for Computational Linguistics.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lakoff, George. 2002. *Moral politics: How liberals and conservatives think*, 2nd edn. Chicago: University of Chicago Press.

Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Landauer, Thomas K. & Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104(2). 211–240.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Stanford: Stanford University Press.

Langacker, Ronald W. 1999. *Grammar and conceptualization*. Berlin; New York: De Gruyter Mouton.

Levy, Omer & Yoav Goldberg. 2014. Neural word embedding as implicit matrix factorization. *Advances in Neural Information Processing Systems* 27. 2177–2185.

Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang & Martin A. Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449(7163). 713–716.

Louwerse, Max M. 2007. Symbolic or embodied representations: A case for symbol interdependency. In Thomas K. Landauer, Danielle S. McNamara, Simon Dennis & Walter Kintsch (eds.), *Handbook of latent semantic analysis*, 107–120. Mahwah: Erlbaum.

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

Lynott, Dermott, Louise Connell, Marc Brysbaert, James Brand & James Carney. 2020. The Lancaster Sensorimotor Norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behaviour Research Methods* 52. 1271–1291.

McCarthy, Diana, Marianna Apidianaki & Katrin Erk. 2016. Word sense clustering and clusterability. *Computational Linguistics* 42(2). 245–275.

Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. https://doi.org/10.48550/arXiv.1301.3781.

Miller, Joanne L. 1997. Internal structure of phonetic categories. *Language and Cognitive Processes* 12(5–6). 865–870.

Mitra, Sunny, Ritwik Mitra, Martin Riedl, Chris Biemann, Animesh Mukherjee & Pawan Goyal. 2014. That's sick dude! Automatic identification of word sense change across different timescales. *Proceedings of the 52nd annual meeting of the Association for Computational Linguistics, Volume 1: Long papers*, 1020–1029. Association for Computational Linguistics.

Mitra, Sunny, Ritwik Mitra, Suman Kalyan Maity, Martin Riedl, Chris Biemann, Pawan Goyal & Animesh Mukherjee. 2015. An automatic approach to identify word sense changes in text media across timescales. *Natural Language Engineering* 21(5). 773–798.

Monaghan, Padraic. 2014. Age of acquisition predicts rate of lexical evolution. *Cognition* 133(3). 530–534.

Monaghan, Padraic & Seán G. Roberts. 2019. Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing. *Cognition* 186. 147–158.

Navigli, Roberto & Mirella Lapata. 2009. An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(4). 678–692.

Navigli, Roberto, Paola Velardi & Stefano Faralli. 2011. A graph-based algorithm for inducing lexical taxonomies from scratch. In *Proceedings of the twenty-second international joint conference on artificial intelligence*, *Vol. 3*, 1872–1877. AAAI Press.

Nosofsky, Robert M. 1988a. Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14(4). 700–708.

Nosofsky, Robert M. 1988b. Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14(1). 54–65.

Nygaard, Lynne C., S. Alexandra Burt & Jennifer S. Queen. 2000. Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(5). 1228–1244.

Pagel, Mark, Quentin D. Atkinson & Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449. 717–720.

Paul, Hermann. 1995 [1880]. *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemeyer Verlag.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan L. Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins.

Pierrehumbert, Janet. 2003. Probabilistic phonology: Discrimination and robustness. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probabilistic linguistics*, 177–228. Cambridge, MA: MIT Press.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Plank, Frans. 2010. Variable direction in zero-derivation and the unity of polysemous lexical items. *Word Structure* 3(1). 82–97.

Pulvermüller, Friedemann. 2005. Brain mechanisms linking language and action. *Nature Reviews Neuroscience* 6. 576–582.

R Core Team. 2023. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Ransmayr, Jutta, Karlheinz Mörth & Matej Ďurčo. 2017. AMC (Austrian Media Corpus): Korpusbasierte Forschungen zum österreichischen Deutsch. In Claudia Resch & Wolfgang U. Dressler (eds.), *Digitale Methoden der Korpusforschung in Österreich*, 27–38. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Reijnierse, W. Gudrun, Christian Burgers, Marianna Bolognesi & Tina Krennmayr. 2019. How polysemy affects concreteness ratings: The case of metaphor. *Cognitive Science* 43(8). e12779.

Rodina, Julia, Daria Bakshandaeva, Vadim Fomin, Andrey Kutuzov, Samia Touileb & Erik Velldal. 2019. Measuring diachronic evolution of evaluative adjectives with word embeddings: The case for English, Norwegian, and Russian. In *Proceedings of the 1st international workshop on computational approaches to historical language change*, 202–209. Association for Computational Linguistics.

Rosch, Eleanor. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104(3). 192–233.

Rosch, Eleanor. 1978. Principles of categorization. In Eleanor Rosch & Barbara B. Lloyd (eds.), *Cognition and categorization*, 27–48. Hillsdale: Erlbaum.

Rosch, Eleanor & Carolyn B. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology* 7(4). 573–605.

Rumelhart, David E., James L. McClelland & PDP Research Group. 1986. *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations*. Cambridge, MA: MIT Press.

Schlechtweg, Dominik, Enrique Castaneda, Jonas Kuhn & Sabine Schulte im Walde. 2021a. Modeling sense structure in word usage graphs with the weighted stochastic block model. In *Proceedings of *SEM 2021: The tenth joint conference on lexical and computational semantics*, 241–251. Association for Computational Linguistics.

Schlechtweg, Dominik, Nina Tahmasebi, Simon Hengchen, Haim Dubossarsky & Barbara McGillivray. 2021b. DWUG: A large resource of diachronic word usage graphs in four languages. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, 7079–7091. Association for Computational Linguistics.

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Smolensky, Paul. 1987. Connectionist AI, symbolic AI, and the brain. *Artificial Intelligence Review* 1. 95–109.

Sweetser, Eve. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

Tahmasebi, Nina. 2018. A study on Word2Vec on a historical Swedish newspaper corpus. *Proceedings of the digital humanities in the Nordic countries 3rd conference*, 25–37. Aachen: CEUR-WS.org.

Tahmasebi, Nina & Thomas Risse. 2017. Finding individual word sense changes and their delay in appearance. *Proceedings of the international conference "Recent advances in natural language processing"*, 741–749. Shoumen: INCOMA Ltd.

Traugott, Elizabeth Closs. 2018. Rethinking the role of invited inferencing in change from the perspective of interactional texts. *Open Linguistics* 4(1). 19–34.

Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

Turney, Peter & Saif Mohammad. 2019. The natural selection of words: Finding the features of fitness. *PLoS ONE* 14(1). e0211512.

Uban, Ana, Alina Maria Ciobanu & Liviu P. Dinu. 2019. Studying laws of semantic divergence across languages using cognate sets. In *Proceedings of the 1st international workshop on computational approaches to historical language change*, 161–166. Association for Computational Linguistics.

Ullmann, Stephen. 1962. *Semantics: An introduction to the science of meaning*. Oxford: Blackwell.

Ustalov, Dmitry, Alexander Panchenko & Chris Biemann. 2017. Watset: Automatic induction of synsets from a graph of synonyms. *Proceedings of the 55th annual meeting of the Association for Computational Linguistics, Vol. 1: Long papers*, 1579–1590. Association for Computational Linguistics.

Vigliocco, Gabriella, Lotte Meteyard, Mark Andrews & Stavroula Kousta. 2009. Toward a theory of semantic representation. *Language and Cognition* 1(2). 219–247.

Vylomova, Ekaterina, Sean Murphy & Nicholas Haslam. 2019. Evaluation of semantic change of harm-related concepts in psychology. In *Proceedings of the 1st international workshop on computational approaches to historical language change*, 29–34. Association for Computational Linguistics.

Westermann, Gert, Nicolas Ruh & Kim Plunkett. 2009. Connectionist approaches to language learning. *Linguistics* 47(2). 413–452.

Whaley, Charles P. 1978. Word–nonword classification time. *Journal of Verbal Learning and Verbal Behavior* 17(2). 143–154.

Widdows, Dominic & Beate Dorow. 2002. A graph model for unsupervised lexical acquisition. In *Proceedings of the 19th international conference on computational linguistics*. Association for Computational Linguistics.

Winter, Bodo & Mahesh Srinivasan. 2022. Why is semantic change asymmetric? The role of concreteness and word frequency and metaphor and metonymy. *Metaphor and Symbol* 37(1). 39–54.

Wood, Simon N. 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall/CRC.

Zhou, Kaitlyn, Kawin Ethayarajh, Dallas Card & Dan Jurafsky. 2022. Problems with cosine as a measure of embedding similarity for high frequency words. *Proceedings of the 60th annual meeting of the Association for Computational Linguistics, Vol. 2: Short papers*, 401–423. Association for Computational Linguistics.

Zipf, George K. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Boston: Houghton Mifflin.

Zipf, George K. 1949. *Human behavior and the principle of least effort*. New York: Addison-Wesley.