To appear in Cognition.

Conceptual relations predict colexification across languages

Yang Xu[a,*,1], Khang Duong[b,*], Barbara C. Malt[c], Serena Jiang[d], Mahesh Srinivasan[e]

a Department of Computer Science, Cognitive Science Program, University of Toronto, Toronto, ON M5S 3G8, Canada

b Computational Sciences Program, Minerva Schools at KGI, San Francisco, CA 94103, USA

c Department of Psychology, Lehigh University, Bethlehem, PA 18015, USA

d Computer Science Program, University of California, Berkeley, CA 94720, USA

e Department of Psychology, University of California, Berkeley, CA 94720, USA

∗ Equal contribution

[1] Correspondence: yangxu@cs.toronto.edu

Abstract

In natural language, multiple meanings often share a single word form, a phenomenon known as colexification. Some sets of meanings are more frequently colexified across languages than others, but the source of this variation is not well understood. We propose that cross-linguistic variation in colexification frequency is non-arbitrary and reflects a general principle of cognitive economy: More commonly colexified meanings across languages are those that require less cognitive effort to relate. To evaluate our proposal, we examine patterns of colexification of varying frequency from about 250 languages. We predict these colexification data based on independent measures of conceptual relatedness drawn from large-scale psychological and linguistic resources. Our results show that meanings that are more frequently colexified across these languages tend to be more strongly associated by speakers of English, suggesting that conceptual associativity provides an important constraint on the development of the lexicon. Our work extends research on polysemy and the evolution of word meanings by grounding cross-linguistic regularities in colexification in basic principles of human cognition.

Keywords: Lexicon; Cross-linguistic variation; Colexification; Concepts; Cognitive economy

Conceptual relations predict colexification across languages

1. Introduction

A central property of natural language is the ability to express an unbounded set of ideas using a finite set of words. One source of this expressivity is colexification—the phenomenon in which multiple meanings are expressed by a single word form (Comrie, 1989; Greenberg, 1966; Majid, Jordan, & Dunn, 2015; Wierzbicka, 1996). Examination of colexification across languages reveals interesting variation in how often sets of meanings are expressed by the same word form (François, 2008; Srinivasan & Rabagliati, 2015; Thompson, Roberts, & Lupyan, 2018; Youn et al., 2016). Some patterns of colexification occur frequently across languages (e.g., languages often use a single word form to express the meanings of both "fire" and "flame"), while other patterns occur less often (e.g., fewer languages have a word form that expresses the meanings of both "fire" and "fever"); see Figure 1A. The present study aims to shed light on the origin of this variation.

Why might some patterns of colexification be more common across languages than others? One possibility is that the meanings that are most often colexified across languages are those that are most strongly conceptually related: i.e., colexification may be driven by the structure of semantic memory. This proposal is motivated by the idea that colexification develops in language due to the dual needs for interlocutors to extend existing words to communicate new ideas and for language learners to quickly acquire their language's lexicon (Ramiro, Srinivasan, Malt, & Xu, 2018; Srinivasan & Rabagliati, 2015; Xu, Malt, & Srinivasan, 2017). More conceptually related meanings may be more likely to be colexified, by this account, because their relatedness facilitates the communication and acquisition of new word meanings. For example, it may be easier for speakers to convey new word meanings by re-using existing words that have highly related meanings, compared to by using existing words that carry more weakly related meanings. Further, it may be easier for language learners to acquire words with multiple, highly related meanings than words with less related meanings.

Some minimal degree of relatedness may be needed before meanings are likely to

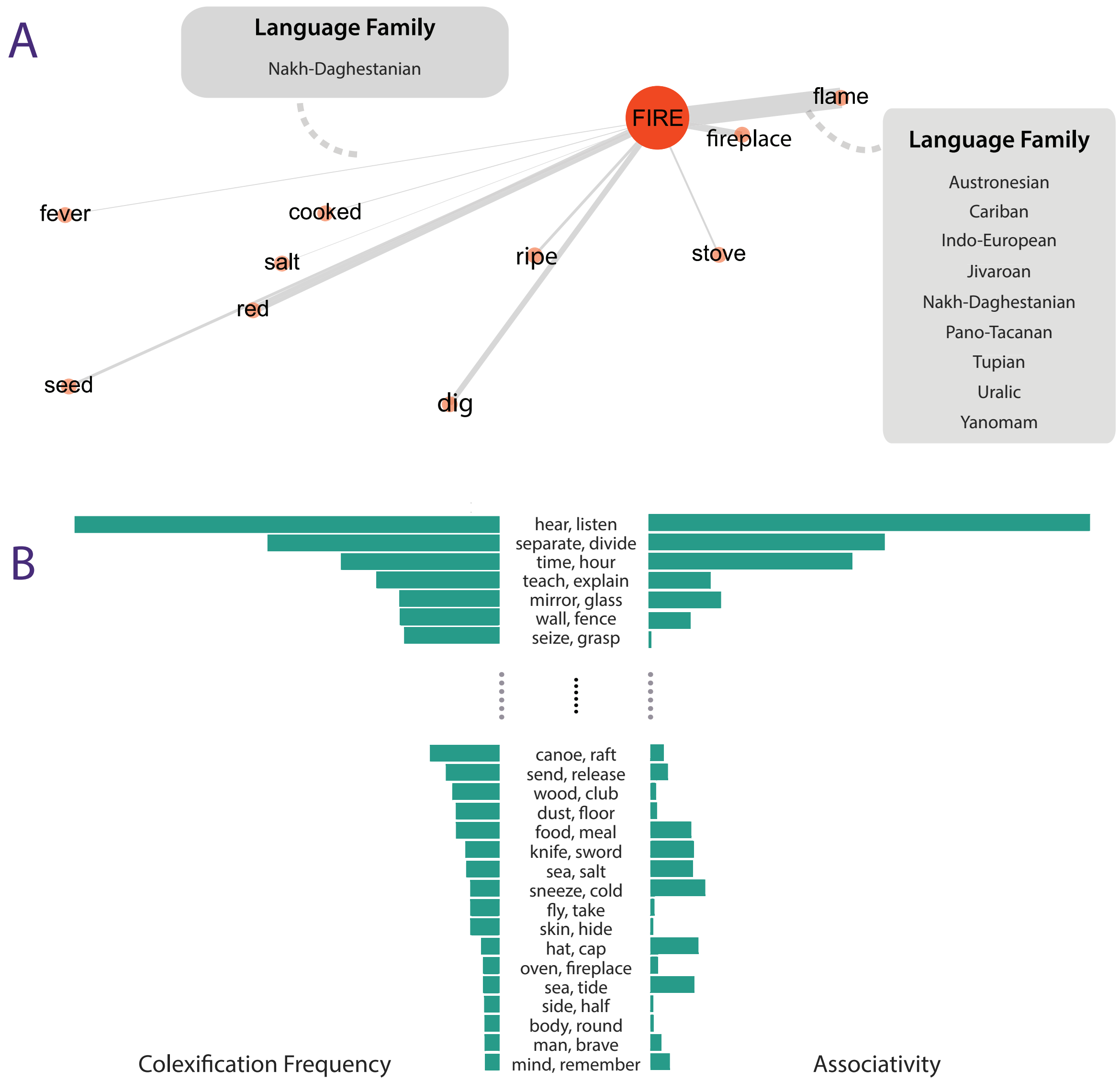

Figure 1. Illustration of colexification across languages and the overall hypothesis. A) A small subset of meanings that colexify with the meaning of “fire”. A gray edge indicates colexification between two meanings; edge width indicates frequency of colexification across languages. Examples of language families in which colexification exists for the meaning pairs “fire, flame” and “fire, fever” are shown. Data are visualized using the tSNE algorithm (Maaten & Hinton, 2008). B) Illustration of the frequency gradient in a subset of colexification across 1310 meanings and 246 languages in the Intercontinental Dictionary Series (Lars, Comrie, & Saxena, 2013), with corresponding degrees of association between meanings measured independently through a game of conceptual association.

be co-lexified; one would not expect to find bicycles and cats, or bananas and snow, to be colexified across many languages. Our proposal, however, goes beyond the idea that relatedness is a prerequisite for the emergence of colexification. We suggest more specifically that the observed *gradient* of colexification probability across languages tracks the *strength* of conceptual relatedness in language users' minds.

This proposal has its roots in prior research. For example, Gentner and Bowerman's Typological Prevalence Hypothesis (Gentner & Bowerman, 2009) proposes that semantic groupings that are more common across languages will be those that are more natural and hence easier for children to learn. Relatedly, other theories have suggested that the function of polysemy and language change is to allow for maximal expressivity with minimal cognitive expenditure (Jespersen, 1959) – a prominent proposal that has been framed in linguistics as the principle of least effort (Geeraerts, 1997; Jespersen, 1959; Zipf, 1949) and in cognitive psychology as the principle of cognitive economy (Rosch, 1978). Such proposals imply that meanings that are more conceptually related may facilitate the communication of new word meanings and are hence more likely to be packaged together.

Providing empirical tests of such proposals is not easy. Gentner and Bowerman (2009) noted that they often employ circular logic: e.g., Children learn X before Y because X is more natural, and we know that X is more natural because children learn it more easily. They broke out of this circularity by examining the relation of order of acquisition to typological prevalence, but they did so for learners of only two languages and for only spatial terms. More recently, Youn and colleagues (2016) analyzed a larger set of colexification frequency data from 4 domains (with 22 seeding concepts) and about 80 languages, creating semantic maps in which pairs of colexified meanings were connected by edges with thickness reflecting frequency in the graph. They found that distinct clusters of co-lexification around core concepts such as "the sun", "water", and "rock" reliably occurred across languages. Youn et al. (2016) suggested that these semantic networks might shed light on a universal conceptual structure shared by speakers of different languages. Because the semantic networks are derived from colexification data, though, they cannot be taken as an explanation of colexification frequency. The problem of circularity arises whenever a pattern of data is used to explain itself.

Previous investigations of colexification thus leave unanswered a key question: Why are some meanings more frequently colexified than others across languages, resulting in gradients of colexification as illustrated in Figure 1A? Although explanations in terms of ease of learning or use are intuitively appealing, they are not the only possible ones. As Gentner and Bowerman (2009) note, the frequency of particular colexifications is likely to be influenced by socio-political factors. Over time, some language families expand through migration and language contact, while others lose speakers. Variation in colexification frequency across languages may arise as a consequence. Frequency might come into play in another way as well: Concepts that are more often used may have more opportunities to become linked to a shared word form. Additionally, Piantadosi et al. (2012) point to ease of disambiguation as a key factor explaining colexification. Under this account, meanings that occur in the most distinct contexts (given some minimal relatedness) may tend to be most often colexified. Such variables could create colexification gradients unrelated to or in contrast to strength of conceptual relatedness.

Other factors may also work against observing a relation between colexification frequency and conceptual relatedness. What constitute the strongest conceptual links may be substantially culture-bound, mitigating any relation of conceptual relatedness to cross-linguistic variation in colexification frequency. At the same time, strongly related ideas such as “sister” and “brother”, or “good” and “evil” which tend to be mentioned in the same contexts, may benefit from being distinguished with separate lexemes rather than being colexified (Kemp & Regier, 2012), and this too may be culture-dependent (e.g., snow and ice may be more important to distinguish for northern cultures; Regier et al., 2016). Last, colexification encompasses cases of homophony, hyperonmy, and multiple forms of polysemy including metaphorical and metonymic meaning relations, bringing into question whether there is an overarching principle that explains cross-linguistic gradients of colexification. In short, empirical tests are needed to determine whether there is a role for conceptual relatedness in explaining colexification frequency across languages.

The goal of the current work is to test the hypothesis that cross-linguistic variation in colexification frequency is non-arbitrary and reflects a tendency in lexical evolution toward cognitive economy (Geeraerts, 1997; Jespersen, 1959; Rosch, 1978; Zipf, 1949). We postulate that meanings that require less cognitive effort to relate (e.g., fire' and flame) should be more likely to enter the lexicons of the world's languages under a shared word form than meanings that are harder to relate (e.g., "fire"→"fever"). Using large-scale data from psychological and linguistic resources, we test whether measures of conceptual relatedness can predict not only which meanings are likely to be colexified across languages, but, further, how often these meanings will be colexified, even after controlling for genetic relatedness among languages. We also examine whether these variables predict cross-linguistic colexification frequency better than other variables that do not directly capture relations between concepts.

Our hypothesis that conceptually related meanings facilitate the communication and acquisition of word meanings is consistent with prior work in psycholinguistics and language development. For example, when speakers intend to express a new meaning for which they do not have a label, existing words with related meanings are particularly accessible and retrievable within the context (Levelt, Roelofs, & Meyer, 1999). Listeners also find it easier to compute a new intended meaning of a word if they can access a highly related, existing word meaning to constrain their interpretation (H. H. Clark & Gerrig, 1983; Klepousniotou, Titone, & Romero, 2008; Rodd et al., 2012), and children find it especially easy to learn a new meaning of an existing word when the new meaning is related to and predictable from previously learned meanings of the word (Srinivasan, Al-Mughairy, Foushee, & Barner, 2017; Srinivasan, Berner, & Rabagliati, in press). These findings follow from more general principles of memory. Models of semantic memory have posited association as a key organizing principle (Anderson & Bower, 1973; Collins & Loftus, 1975; Rogers & McClelland, 2004). In such models, retrieval of one piece of information given a prior one is driven by the strength of association between the two; strongly associated pieces of information will more readily retrieve one another than more weakly associated ones.

Our proposal that more frequently colexified meanings should be more conceptually related is also consistent with recent evidence on how new word meanings enter lexicons over time (Ramiro et al., 2018; Xu et al., 2017). For example, Ramiro et al. (2018) found that over the past 1,000 years, new meanings in English have much more often entered the lexicon by attaching to existing word forms than by creation of new forms. Ramiro and colleagues also found that words tend to gain new related meanings over time via a process of chaining (Bybee, Perkins, & Pagliuca, 1994; Lakoff, 1987; Malt, Sloman, Gennari, Shi, & Wang, 1999; Ramiro et al., 2018; Xu, Regier, & Malt, 2016). In chaining, a novel meaning is linked with existing meanings of a word that are closely related, forming a chain-like structure in semantic space (e.g., “fire”→“flame”→“red”). Chaining allows word meanings to extend in small steps such that the addition of new senses to a word form should require minimal cognitive effort to generate, use, and learn. If chaining is the main strategy through which words have developed new meanings over history, we should expect more frequently colexified meanings across different languages to be more strongly related due to this process of word meaning extension.

Finally, our proposal extends work suggesting that the lexicon has been shaped by the pressure to support efficient communication while minimizing ambiguity (Kemp & Regier, 2012; Kirby, Tamariz, Cornish, & Smith, 2015; Piantadosi, Tily, & Gibson, 2012; Regier, Carstensen, & Kemp, 2016). For example, prior research suggests that short and frequent words often carry multiple meanings because these words are easy for speakers to use and the ambiguity created can typically be readily resolved from context (Piantadosi et al., 2012). Our research extends this previous work by asking whether sets of meanings that are more conceptually related are more likely to be grouped under a common label across languages.

2. Materials and methods

2.1 Cross-linguistic data of colexification

We drew on one of the largest databases of digitized lexicons in terms of the number of semantic domains and languages represented: The Intercontinental Dictionary Series (IDS) (Lars et al., 2013). The IDS contains 1310 entries of meaning that are organized into 22 semantic domains such as "the physical world", "sense perception" and "religion and belief". They are identified by one (e.g., *world*) or a few (e.g., *earth*, *land*) English words and explicated in more detail in linked entries (e.g., for *world*, *The Earth with all its inhabitants and all the things upon it*). The IDS word senses were designed to represent meanings that are common to diverse languages, and the set of meanings from IDS was adapted from Carl Buck's dictionary (Buck, 1949) that originally focused on comparative work of lexicons in Indo-European languages. Later extension of this database included meanings from a substantial number of non-Indo-European (and/or underrepresented) languages. The meaning list for each language was initially collected by field researchers and then corrected and maintained by linguistic experts.

We worked with data from 246 languages that have uniquely identifiable ISO codes, which represented 41 language families, 5 climate categories, and 5 geographical regions (see *Supplementary Material* for details). For analytic purposes we defined colexification as cases where two different meanings share the same word form in a given language. This definition of colexification is necessarily open to either polysemy or homophony/homography. We expect that cases of colexification that occur regularly across languages are likely to correspond to polysemy, and those that are rare are more likely to be cases of homophony. We aim to identify whether positioning of a particular colexification on this frequency gradient (which may represent a cline from homophony to polysemy) is related to how conceptually related those meanings are.

One limitation of the IDS is that English is used as the reference language for annotating meanings in the database, even though these annotations are intended to represent word senses found across languages (p 286, Lars, Comrie, & Saxena (2013)).

Our analyses focus on concepts expressible in single word forms in English and are limited at the level of granularity provided by the IDS dictionary (see *Materials and Methods*). This means that we would not be able to detect colexification cases involving concepts that are not expressible by single English words. This potential selection bias restricts the set of possible colexifications that we can examine, but the basic problem we pose holds within the set of available concepts: What drives the gradient of colexification frequency across languages? For instance, it is logically possible that the concepts "moon" and "month" could be rarely colexified, moderately often colexified, or frequently colexified across the IDS languages. Thus, our hypothesis–that concepts that are more strongly related will be more likely to be colexified across languages–is falsifiable even within the incomplete set of concepts that we investigate.

We estimated the frequency of a colexified pair of meanings across languages via stratified bootstrapping (Bickel & Freedman, 1984), which controls for variation in language families, climate categories, and geographical regions. The ensemble of colexified meaning pairs $S = \{(s_i, s_j) : ij \in C\}$ (C stands for the set of indices of colexified pairs) and their frequencies $F = \{f_c(s_i, s_j) : s_i s_j \in S\}$ constitute the colexification data that we sought to explain (see Figure 1B for illustration). The *Supplementary Material* provides details on the procedures and the robustness of the colexification frequencies under the three controlled factors.

Although the IDS project is a pioneering effort that contributes to preserving information on little-known and "non-prestigious" languages (p. 286, Lars, Comrie, & Saxena, 2013), it draws on a subset of the world's language families (e.g., it does not currently include Australian and non-Austronesian languages of Papua New Guinea) and has uneven representations from the language families (e.g., Indo-European is more represented than other language families). For this reason, we also performed analyses based on Monte Carlo simulation (Everett, Blasi, & Roberts, 2015) and found that our results are robust under this procedure. In the Monte Carlo simulation, we only sampled one language per language family, such that we treated each family equally regardless of its representation in IDS. We describe the details of these analyses in

Supplementary Material.

The raw data from IDS are publicly available at `https://ids.clld.org/`. Code for analyses is available under the project name "ids_colexification" at `https://osf.io/d5txn/`.

2.2 Predictors of colexification frequency

We considered five variables to account for cross-linguistic colexification frequencies. Two of these variables directly capture relations (or relatedness) between concepts, and we derived these measures from behavioral measures of conceptual associativity and corpus-based measures of conceptual similarity, respectively. We consider both of these measures as proxies of conceptual relatedness, and as we show, they should be correlated. The three remaining variables (usage frequency and two variables related to metaphoricity) are qualitatively different in that they do not directly capture relations between concepts. They serve as baseline comparison measures capturing complementary influences on colexification. Our methods allow us to control for language genetic relatedness, assessing our variables' explanatory power independent of relatedness.

1. Relatedness by associativity. We considered the first variable of conceptual relatedness as associative strength or associativity, derived from a task in which one or a set of seed words is presented to a participant who responds with the first word (or more) that comes to mind (De Deyne & Storms, 2008). Associative strength has been successfully used in psychology to predict performance in a wide range of tasks including word recognition, similarity judgments, and free and cued recall (De Deyne & Storms, 2008). We defined associativity $a(s_i, s_j)$ as the expected probability with which a person will associate one meaning s_i with another meaning s_j. Our hypothesis predicts that frequently colexified meanings should tend to be highly associated, requiring low cognitive effort, and conversely, less frequently colexified concepts should tend to be less strongly associated (Figure 1B illustrates this hypothesis):

$$f_c(s_i, s_j) \propto a(s_i, s_j) \tag{1}$$

We consider a linear relation between frequency and associativity because it offers a simple, parameter-free formulation, but we do not rule out the possibility of complex, non-linear relations. We used data from a large-scale online association game, Human Brain Cloud (abbreviated as HBC), to estimate the associativity of meanings in the colexification network, and we also tested our proposal using a separate, smaller dataset from a controlled word association experiment from University of South Florida (abbreviated as USF) (Nelson, McEvoy, & Schreiber, 1998). More details of our measures of conceptual association are provided in the *Supplementary Material*.

2. Relatedness by similarity. We considered a second proxy of conceptual relatedness via similarity. The similarity model postulates that colexification depends on the degree of similarity between meanings. Similarity typically refers to overlap in meaning. Although similarity may be operationalized in different ways such as feature lists, spatial representations, or structured representations (see Markman, 1999; Medin et al.,1993), we considered more scalable vector-based models of similarity based on natural language usage (i.e., words with similar linguistic context are considered to be similar under this model). These are known to correlate with human similarity judgments (Pereira, Gershman, Ritter, & Botvinick, 2016), although they are also influenced by word associations (Estes & Simmons, 2006).

In particular, we used word2vec embedding (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) that has been shown to be a useful way to approximate similarity (Pereira et al., 2016). We derived these similarities from large-scale corpus data by making use of pre-trained word embeddings from word2vec [1]. For a given pair of concepts, we operationalized similarity by taking the standard cosine distance between word embeddings of these concepts.

The similarity model predicts that concepts that are used in more similar linguistic contexts should be more frequently colexified across languages. Prior research

[1] `https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTTlSS21pQmM/edit`

suggests that although the meanings of polysemous words are semantically related, they can often be dissimilar (Klein & Murphy, 2001, 2002; Rabagliati, Marcus, & Pylkkänen, 2011). Overall, we expected that our models of similarity and associativity would better account for the cross-linguistic colexification frequency distribution than models that do not directly capture conceptual relations.

3. Baseline of usage frequency. As noted before, concepts that are more often used may have more opportunities to become linked to a shared word form. To account for this, we considered a frequency-based model postulating that colexification is independent of intrinsic relations between meanings, but instead depends on the expected usage frequency of encountering each of those meanings. As a proxy for frequency of a concept in thought and communication, we obtained frequencies from the Google Book English corpus (Michel et al., 2011) [2], where we estimated them from the year of 1949—anchored to the publication date for Carl Buck's dictionary (Buck, 1949) that the IDS concepts are based on. For a given pair of concepts, we took the average frequency of words that define those concepts.

4 & 5. Baselines of Metaphoricity. Abstract concepts are often understood by reference to more concrete concepts (e.g., an argument as a war) (Lakoff & Johnson, 1980), and words that first arise with concrete senses often later gain an abstract one (Xu et al., 2017). We considered two versions of a metaphor-based model to take into account the possibility that colexification is driven by metaphorical tendency between concepts. We measured potential metaphoricity by contrasts in human judgments on the dimensions of concreteness and of emotional valence (i.e., positive or negative emotion) for each meaning. For instance, *slime* earlier referred to "semi-fluidity"–a relatively concrete and emotionally-neutral concept, but was later metaphorically extended to denote "moral evil"–a more abstract and negatively valenced concept. We take metaphoricity to be high if there is a large contrast between concepts in either the concreteness or valence dimensions. These models predict that concepts that differ in concreteness or valence would tend to be more frequently

[2] `http://storage.googleapis.com/books/ngrams/books/datasetsv2.html`

colexified across languages. Previous work has shown that these dimensions have independent predictive value in accounting for historical metaphorical extensions of meaning (Xu et al., 2017), and we expect metaphor to be responsible for driving some colexification across languages (as in the examples above). For a pair of concepts, we defined the concreteness model by considering the contrast in concreteness of the two concepts by taking the absolute difference: $|concreteness(s_i) - concreteness(s_j)|$. Similarly, we defined the valence model by considering the contrast in valence of the two concepts. We obtained human ratings of concreteness and valence from large psychological norms published in Brysbaert, Warriner, and Kuperman (2014) and Warriner, Kuperman, and Brysbaert (2013). A limitation of these measures of metaphoricity is that they do not directly capture relations between concepts (obtaining better measures of metaphor would require a deep understanding of structural similarities between conceptual domains). Rather, the models we considered here predict that any two meanings that differ in valence or concreteness will have an increased likelihood of colexification, even if they might be conceptually unrelated. We were interested in whether this measure of potential metaphoricity might be complementary with our relation-based measures described above.

For all the measures, we assumed equivalence between an English word in the measurement data (e.g., “fire”) and an IDS meaning (e.g., fire). The assumption is made of necessity because measures directly related to word senses are not available at a broad scale. This assumption limits the model performances because word-based measures of meaning are likely to be imprecise, given that many words do encode multiple related meanings. By using word association and other alternative lexical measures derived from English, our analyses focus on explaining relations among colexified IDS meanings that can be labelled by individual English word forms as opposed to phrases. Despite this fact, our analyses draw on a large subset (1,080 out of 1,310, i.e., 82.4%) of the senses recorded in the IDS.

If more fine-grained measures were available, predictive power could increase. For instance, since “sheet” in English encompasses both bed sheets and sheets of paper, then if “pillow” is the main associate of “sheet”, we might miss predicting the colexification of sheets of paper and envelopes in some languages. Our data thus provide lower bounds on the predictive power of our measures.

3. Results

We performed analyses in three steps. First, we show that the variables described above—particularly those that directly capture conceptual relations—can distinguish between colexification patterns that are attested in natural languages and those that are unattested. Second, we show that among the set of variables, association best explains why some concepts are more frequently colexified across languages than others. Finally, we show that the degree of associativity between concepts—i.e., whether concepts are directly associated or indirectly associated via intervening concepts—recapitulates the gradient of colexification frequency across our large sample of languages.

The first analysis provides a sanity check on the predictive value of the proposed variables on colexification patterns. Specifically, we examined whether the five variables we have described can predict attested patterns of colexification at all, without taking into account the cross-linguistic frequency of these patterns of colexification. For any pair of concepts in the IDS database, our models predicted whether it should be colexified or not in any of the languages from the database. An ideal model should predict that concept pairs such as “fire” and “flame” will be colexified in a language, whereas pairs such as “fire” and “grandfather” should not be colexified.

To evaluate the variables, we performed a standard logistic regression analysis on each individual variable and a multiple logistic regression analysis for all the variables combined. For each of these analyses, we then split the data into two groups: 1) Attested, defined as pairs of concepts that have been attested to be colexified in at least two of the set of IDS languages we considered—this criterion sets the minimal requirement (or a lower bound) on inter-language agreement on the colexified concept pairs; 2) Null, defined as pairs of concepts that have been attested as colexified for zero or one language in the set. We cast our analyses as a binary classification problem, where we used logistic regression to determine whether the variables of interest can successfully predict the group membership of any queried pair of concepts. We applied a standard (10-fold) cross validation in these analyses, and we report model accuracy only from the held-out predictive set (i.e., $\frac{1}{10}th$ of data) while using the remaining set (i.e., $\frac{9}{10}th$ of data) to estimate the logistic models. Because there are many more pairs in the null group ($N = 574,785$) than in the attested group ($N = 8,909$), we balanced the data between the two groups by sampling from the null group and equating the sample size to that of the attested group. We report the cross-validated predictive performance based on mean accuracy from 20 sampled sets of the null group and the fixed attested group (these sets are identical for all variables we considered).

Figure 2A summarizes the results. Overall, variables that capture conceptual relations, including associativity and similarity, performed well above chance (50%). In particular, the word association resources we worked with predict $\sim 70\%$ of the patterns in the colexification data, with false positive rates 0.6% [95% CI: 0.5,0.7, from bootstrapping] (for HBC) and 0.4% [95% CI: 0.3,0.6] (for USF), and the false negative rates 32.7% [95% CI: 32.3,33.0] (for HBC) and 33.6% [95% CI: 33.0,34.2] (for USF). These results indicate that word association has high precision in predicting the colexification patterns and lower recall (see an analysis later that improves recall by considering word association at higher orders). On the other hand, the remaining variables that do not directly capture conceptual relations showed significantly worse predictive performance ($p < 0.001$, permutation test with 1000 shuffles). Interestingly, models based on metaphoricity predicted colexification above chance (Concreteness: $t(99) = 277.498$, Valence: $t(99) = 197.487$; one-tail $p < 0.001$ in both cases), suggesting that metaphoricity plays a role in colexification. However, as noted above, the models based on metaphoricity had substantially less predictive power than those based on associativity or similarity. Usage frequency performed at chance level, suggesting that the individual frequency of each concept within a pair plays little role in explaining whether that pair of concepts will be colexified. Figure 2B shows that the two measures of conceptual relatedness are the most correlated variables among the five variables we have considered (Spearman's $\rho = 0.48$, $p < 0.001$). The remaining variables show substantially less degrees of correlation (Spearman's correlations among all variable pairs are significant at $p < 0.001$ except for Associativity vs Frequency with $p = 0.80$ and Valence vs Frequency with $p = 0.17$), suggesting that these variables contain orthogonal information.

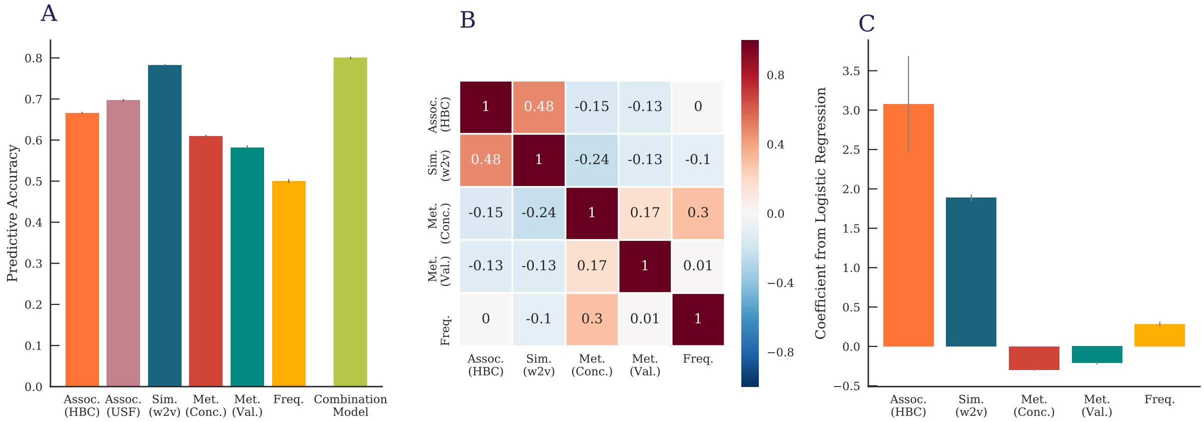

Figure 2. Results from predictive analyses of cross-linguistic colexification. A) Cross-validated predictive accuracies in classifying colexified versus non-colexified sense pairs from individual variables and variables in combination, using logistic regression. B) Spearman correlations among the five predictive variables of colexification. C) Logistic regression coefficients of different variables from the combination model. Abbreviations are used for the following variables: "Assoc"→"Associativity", "Sim"→"Similarity", "Met. (Conc.)"→"Metaphoricity (Concreteness)", "Met. (Val.)"→"Metaphoricity (Valence)", "Freq."→"Frequency". Error bars indicate deviations in predictive accuracy from the randomized cross validation sets.

To take into account the relative contributions of all the variables, we performed a standard multiple logistic regression analysis. Figure 2A shows that the predictive accuracy of the combined model is on par with the similarity model. Importantly, by examining the variable coefficients (see Figure 2C) we found that associativity and similarity have the highest coefficients among the five variables of interest. In particular, coefficients of metaphoricity and usage frequency are close to zero in this controlled analysis, while the coefficients of the associativity and similarity variables are both significantly above zero ($p < 0.001$ from permutation tests with 1000 shuffles). We provide examples of the most frequently colexified concepts, as well as examples of the most similar concepts and most strongly associated concepts in the *Supplementary Material*. In sum, these initial results suggest that conceptual relations are central to predicting whether a pair of concepts will be colexified in a language.

The second set of analyses go beyond binary predictions of whether a colexification pattern will be attested in natural language or not. Specifically, we evaluated how well each variable and their combination account for cross-linguistic variability in how frequently pairs of concepts are colexified. Because the estimation of cross-linguistic colexification frequency is potentially influenced by factors such as language family, climate, and geography (e.g., languages within the same language family are likely to show similar colexification patterns, which might contribute to over-counting in frequency estimation), we compared the variables by controlling explicitly for each of these external factors. Concretely, we correlated these variables against colexification frequencies for each of the controlled external factors, which are based on 1,000 bootstrapped samples (see *Supplementary Material* for the construction and robustness of the colexification frequencies). We expect the dominant variable to best correlate with colexification frequencies under all three controlled factors.

Figure 3 summarizes the results. Among all variables of interest, we found that associativity best correlates with cross-linguistic frequency of colexification under the three controlled cases. In particular, the associativity variable significantly outperformed the second best variable, similarity ($p < 0.001$ for all cases, permutation test with 1000 shuffles), and both of these variables performed substantially better than the remaining three baseline variables. In the combined linear regression model, we found that the overall correlation between these variables combined and colexification frequency is no greater than that from associativity alone, and that the coefficients of the five variables consistently support the dominant role of associativity in regression against colexification frequency. Together, these results show that colexification frequency is best accounted for by models of conceptual relatedness and particularly conceptual associativity, supporting the idea that concepts that are more frequently colexified across languages will tend to be more strongly associated or related.

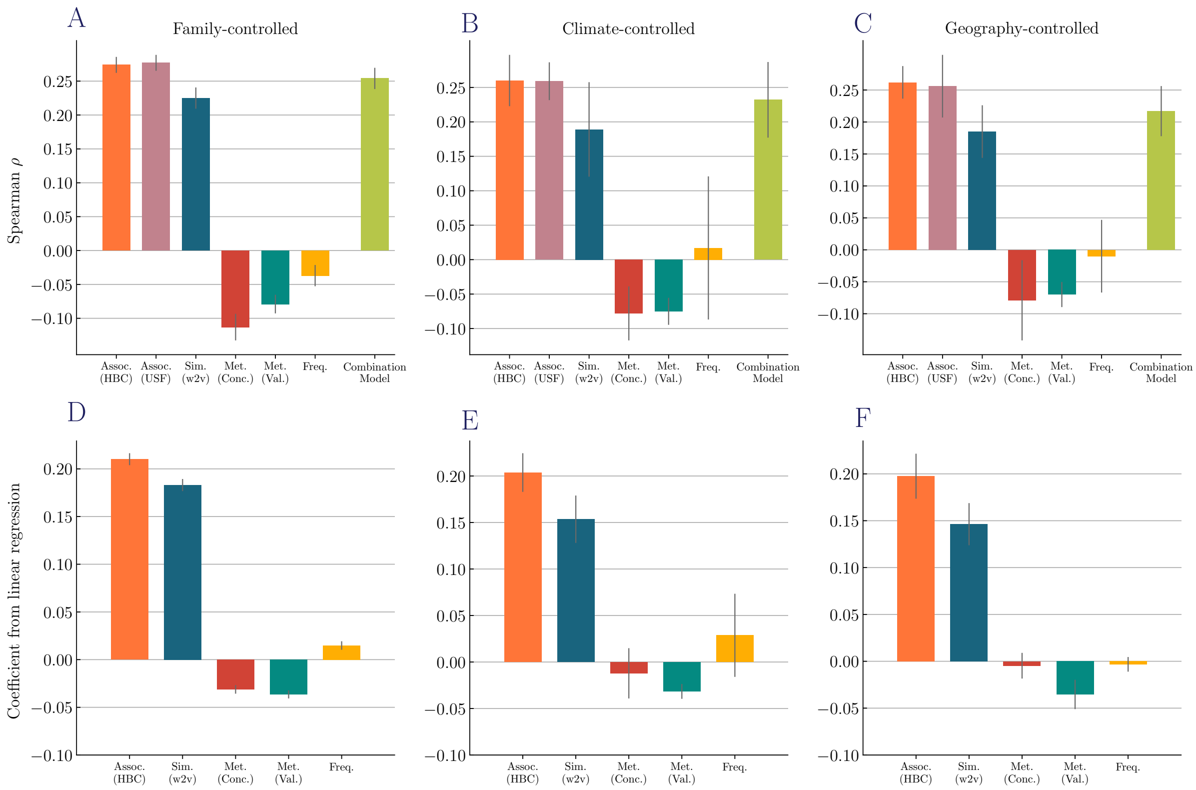

Figure 3. Results of variable correlations with colexification frequencies across languages. Upper panel (A to C) shows results from correlations with individual variables and linear regression from variables in combination. Lower panel (D to F) shows coefficients of individual variables from the linear regression. Colexification frequencies are calculated by controlling for language family, climate category, and geographical region. Abbreviations are used for the following variables: "Assoc"→"Associativity", "Sim"→"Similarity", "Met. (Conc.)"→"Metaphoricity (Concreteness)", "Met. (Val.)"→"Metaphoricity (Valence)", "Freq."→"Frequency". Error bars indicate 95% confidence intervals from bootstrapping.

The analyses described so far have suggested that the direct association of two concepts may be predictive of their colexification frequency across languages. However, it is conceivable that colexification may also coalesce concepts that rely on indirect association, due to an historical process of meaning extension (Ramiro et al., 2018). For example, “birch” and “sap” may be associated via a path that relies on the intermediate concept of “tree”: “birch”→“tree”→“sap”, i.e., a second-order association. Similarly, “nettle” and “water” may be related via a path that requires association of a higher order: “nettle”→“tree”→“world”→“water”. To account for these possibilities, our third set of analyses examined the relationship between the orders of association of two concepts and the cross-linguistic frequency of their colexification. We expected that indirect associations might account for some cases of colexification, but also that concepts that are more directly associated should tend to be more frequently colexified, compared to concepts that are linked only through intermediary associates. More specifically, we predict tiering in the expected colexification frequency of two concepts to recapitulate the order of association between those concepts:

$$E[f_c(s_i, s_j|a(\cdot,\cdot))] \gtrapprox E[f_c(s_i, s_j|a(\cdot,\cdot|s_k))] \gtrapprox E[f_c(s_i, s_j|a(\cdot,\cdot|s_k, s_l))] \quad (2)$$

Here s stands for concepts, and $a(\cdot,\cdot)$, $a(\cdot,\cdot|s_k)$, and $a(\cdot,\cdot|s_k, s_l)$ represent association of first, second and third orders (i.e., associative paths that depend on either zero, one intermediary concept symbolized by s_k, or two intermediary concepts symbolized by s_k and s_l, respectively). Equation 2 says that the gradient of expected frequency of colexification across languages should reflect the tiering or order of association required to relate a pair of meanings, such that meaning pairs that are more remotely associated should be less likely to enter the lexicon, compared to those that are more directly associated.

To test this prediction, we classified the colexified concept pairs into three groups: directly associated, associated under second-order paths, and associated under third-order paths. These three tiers of association provided a coverage of 99.84% (from the HBC data set) and 84.19% (from the smaller USF data set) of the colexification data, and we describe the procedures of this classification and derivations of the ordered association in the *Supplementary Material*. Within each of these groups, we calculated the mean colexification frequencies between the concept pairs.

Figure 4 summarizes the results based on the HBC association data.

Supplementary Material includes similar results for the USF association data. The mean frequency of colexification shows a monotonic decrease as the order of association increases, and this trend is persistent under the controlled variables of language family, climate category, and geographical region. In particular, pairwise comparisons of mean colexification frequency between consecutive association groups reveals significant tiering (HBC: family, 1st vs. 2nd order $t(999) = 765.203$, 2nd vs. 3rd order $t(999) = 319.546$; climate, 1st vs. 2nd order $t(999) = 345.568$, 2nd vs. 3rd order $t(999) = 41.280$; geographical region, 1st vs. 2nd order $t(999) = 280.024$, 2nd vs. 3rd order $t(999) = 86.602$; $p < 0.001$ in all cases; USF: family, 1st vs. 2nd order $t(999) = 628.874$, 2nd vs. 3rd order $t(999) = 445.105$; climate, 1st vs. 2nd order $t(999) = 324.283$, 2nd vs. 3rd order $t(999) = 133.601$; geography, 1st vs. 2nd order $t(999) = 237.572$, 2nd vs. 3rd order $t(999) = 132.769$; one-tail $p < 0.001$ in all cases). These results provide support for the idea that indirect as well as direct associations contribute to colexification while also showing that the cross-linguistic gradient of colexification frequency is recapitulated by the ordering of conceptual association.

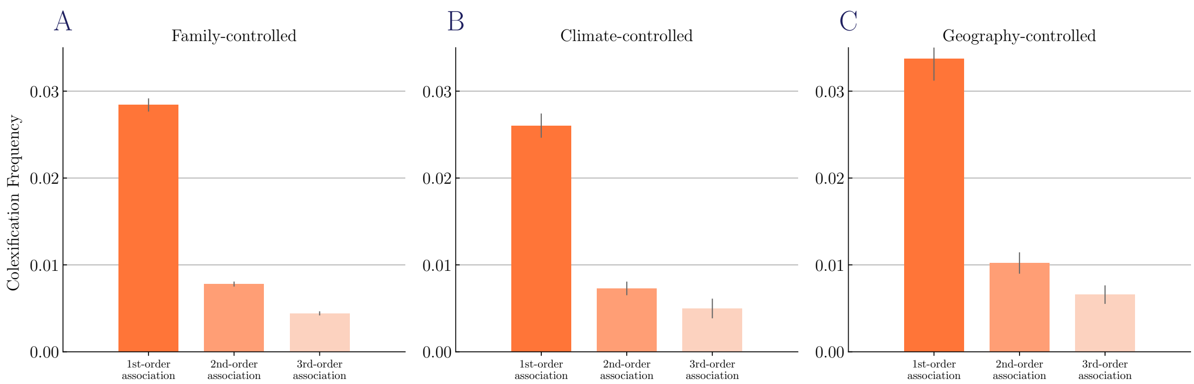

Figure 4. Gradient of colexification frequencies in ordered association sets. Results shown are based on the Human Brain Cloud association data, controlled for language family, climate category, and geographical region. Error bars indicate 95% confidence intervals from bootstrapping.

4. Discussion

Colexification is a fundamental phenomenon contributing to the expressivity of language under a finite lexicon. We have shown that frequency gradients in colexification across a large sample of languages are best predicted by an independent assessment of conceptual association strength. In doing so, we establish the importance of conceptual relations in colexification processes. Our results suggest that although languages vary in how they package meanings into word forms, this variation nevertheless reflects a tendency to reduce cognitive effort of association. As we proposed earlier, it may be relatively easy for speakers to communicate new meanings by re-using existing words with highly associated meanings, and it may be relatively easy for children to learn new word meanings when they can use a highly associated, already-learned word meaning to guide their interpretation. These pressures on language to support the communication and learning of new word meanings may help explain why more frequently colexified meanings across languages tend to be those that are more strongly associable.

This work provides an account for why languages exhibit the gradient of colexification patterns that they do, extending previous descriptive generalizations of cross-linguistic lexical data represented by semantic maps (François, 2008; Youn et al., 2016), and doing so at a large scale. It rules out the possibility that colexification gradients are driven solely by the spread of influence of some languages over others. Even though many different forms of colexification exist (e.g., metaphor, metonymy, homophony), and colexification may be influenced by factors such as the communicative need to distinguish highly related concepts with separate lexemes, or cultural differences in conceptual relations, our findings suggest that conceptual relatedness–and conceptual associativity in particular–places a key constraint on colexification processes.

Our work has implications for the design principles of the lexicon. It connects with a growing line of research which suggests that lexical ambiguity is a necessary property of the lexicon that facilitates efficient communication by allowing readily produced and understood word forms to be re-used (Piantadosi et al., 2012). In particular, existing work has shown that the potential ambiguity that results from colexification can often be resolved from linguistic context, posing no comprehension difficulty (Piantadosi et al., 2012). Our research adds to this prior work by elucidating the design principle that explains why some meanings are more likely to be colexified under a common word form than others. We suggest that certain groupings of meaning under a word form are preferred to others, with the basis of this preference determined in part by how predictable meanings are with respect to each other, measured through strength of association. Indeed, an emphasis on the role of distinctive contexts in permitting colexification might predict that less related meanings would be more likely to be colexified, since such meanings are more likely to arise in distinct contexts. Thus, although the evolution of word meanings may be constrained by the pressure to resolve ambiguity, our findings suggest that it may also shaped by a need for cognitive efficiency (Ramiro et al., 2018; Xu et al., 2017), which has previously been demonstrated primarily in English. Our results indicate that the product of lexical evolution—examined across languages—also reflects the cognitive economy that is seen in diachronic language change.

Our research provides an important avenue into understanding the diversity of lexicalization strategies across languages, but it must ultimately be understood within a broader context. Association may provide the potential for meanings to be colexified, but we expect that other factors, including ones we have already mentioned, will provide constraints on which meanings are actually colexified. For instance, not only would it be unlikely for languages from colder climates to colexify “snow” and “ice”, but it would be unlikely for any language to collapse the meanings of “mother” and “father” under the same word form (Kemp & Regier, 2012; Kemp, Xu, & Regier, 2018) because there is a high need to distinguish these meanings, and they often would not be effectively disambiguated by context. Furthermore, a wide range of cognitive strategies may be employed to extend words to new meanings (Sweetser, 1991; Traugott & Dasher, 2001) or adopt new words (Monaghan & Roberts, 2019), although some strategies might be more prevalent than others due to considerations of cognitive effort and expressivity (Jespersen, 1959). Similarly, work in child language development has documented different strategies children might use in extending their vocabularies to novel situations (E. V. Clark, 1995). A fuller account of lexicalization patterns will also require considering perceptual constraints and the structure of experience (Evans & Levinson, 2009; Malt & Majid, 2013) along with variables such as cultural communicative needs. The present study contributes toward unraveling the nature of lexicalization in natural language.

Acknowledgements

We dedicate this work to Sieghard Beller. We would like to thank Andrea Bender, Simon De Deyne, Gert Storms, Aida Nematzadeh, and three anonymous reviewers for their constructive comments on the manuscript. We also thank Kyle Gabler for sharing the online word association data and Joshua Abbott for helping with data processing. MS is funded by National Science Foundation (SBE-16302040) and a grant from the James S. McDonnell Foundation. YX is funded through an NSERC Discovery Grant, a SSHRC Insight Grant, and a Connaught New Researcher Award.

References

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* Psychology Press.

Bickel, P. J., & Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics*, 470–482.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911.

Buck, C. D. (1949). *A dictionary of selected synonyms in the principal Indo-European languages.* Chicago: University of Chicago Press.

Bybee, J. L., Perkins, R. D., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world.* Chicago: University of Chicago Press.

Clark, E. V. (1995). *The lexicon in acquisition.* Cambridge: Cambridge University Press.

Clark, H. H., & Gerrig, R. J. (1983). Understanding old words with new meanings. *Journal of Verbal Learning and Verbal Behavior*, *22*(5), 591–608.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology.* Chicago: University of Chicago Press.

De Deyne, S., & Storms, G. (2008). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*, 213–231.

Estes, Z., & Simmons, S. (2006). Using latent semantic analysis to estimate similarity. In *Proceedings of the annual meeting of the cognitive science society.*

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–448.

Everett, C., Blasi, D. E., & Roberts, S. G. (2015). Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences*, *112*, 1322–1327.

François, A. (2008). Semantic maps and the typology of colexication: Intertwining polysemy networks across languages. In M. Vanhove (Ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (pp. 163–215). Amsterdam: John Benjamins.

Geeraerts, D. (1997). *Diachronic prototype semantics.* Clarendon Press, Oxford.

Gentner, D., & Bowerman, M. (2009). Why some spatial semantic categories are harder to learn than others: The typological prevalence hypothesis. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Ozcaliskan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 465–480).

Greenberg, J. H. (1966). *Language universals, with special reference to feature hierarchies, janua linguarum, series minor 59.* Mouton: The Hague.

Jespersen, O. (1959). *Language: Its nature, development and origin.* London: Allen & Unwin.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*, 1049–1054.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*, 109–128.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, *45*, 259–282.

Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: Conceptual relations of polysemous senses. *Journal of Memory and Language*, *47*, 548–570.

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *34* (6), 1534.

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

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* University of Chicago press.

Lars, B., Comrie, B., & Saxena, A. (2013). The intercontinental dictionary series: A rich and principled database for language comparison. In B. Lars & A. Saxena (Eds.), *Approaches to measuring linguistic differences* (pp. 285–302).

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22* , 1–38.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9* , 2579–2605.

Majid, A., Jordan, F., & Dunn, M. (2015). Semantic systems in closely related languages. *Language Sciences*, *49* , 1–18.

Malt, B. C., & Majid, A. (2013). How thought is mapped into words. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4* , 583–597.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40* , 230–262.

Markman, A. B. (1999). *Knowledge representation.* Mahwah, NJ: Lawrence Erlbaum Associates.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100* (2), 254–278.

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

Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *Workshop papers in International Conference on Machine Learning.*

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems.*

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

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms.*

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

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*, 280–291.

Rabagliati, H., Marcus, G. F., & Pylkkänen, L. (2011). Rules, radical pragmatics and restrictions on regular polysemy. *Journal of Semantics*, *28*, 485–512.

Ramiro, C., Srinivasan, M., Malt, B. C., & Xu, Y. (2018). Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences*, 2323–2328. doi: 10.1073/pnas.1714730115

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PLoS One*, *11*(4), e0151138.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition*, *40*(7), 1095–1108.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* MIT Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale: Lawrence Erlbaum.

Srinivasan, M., Al-Mughairy, S., Foushee, R., & Barner, D. (2017). Learning language from within: Children use semantic generalizations to infer word meanings. *Cognition*, *159*, 11–24.

Srinivasan, M., Berner, C., & Rabagliati, H. (in press). Children use polysemy to structure new word meanings. *Journal of Experimental Psychology: General*, *https://doi.org/10.31234/osf.io/sr9hq*.

Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua*, *157*, 124–152.

Sweetser, E. (1991). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure.* Cambridge: Cambridge University Press.

Thompson, B., Roberts, S., & Lupyan, G. (2018). Quantifying semantic similarity across languages. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

Traugott, E. C., & Dasher, R. B. (2001). *Regularity in semantic change.* Cambridge: Cambridge University Press.

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

Wierzbicka, A. (1996). *Semantics: Primes and universals.* Oxford: Oxford University Press.

Xu, Y., Malt, B. C., & Srinivasan, M. (2017). Evolution of word meanings through metaphorical mapping: Systematicity over the past millennium. *Cognitive Psychology*, *96*, 41–53.

Xu, Y., Regier, T., & Malt, B. C. (2016). Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, *40*, 2081–2094.

Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., … Bhattacharya, T. (2016). On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences*, *113*, 1766–1771.

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology.* Cambridge: Addison-Wesley.