Title
Understanding the Frequency of a Word by its Associates: A Network Perspective

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Liu, Qiawen
De Deyne, Simon
Jiang, Xiaoya
et al.

Publication Date
2023

Peer reviewed

Understanding the Frequency of a Word by its Associates: A Network Perspective

Qiawen Liu (qliu295@wisc.edu)
Department of Psychology
University of Wisconsin-Madison

Simon De Deyne (simon.dedeyne@unimelb.edu.au)
School of Psychological Sciences
University of Melbourne

Xiaoya Jiang (xjiang267@wisc.edu)
School of Journalism and Mass Communication
University of Wisconsin-Madison

Gary Lupyan (lupyan@wisc.edu)
Department of Psychology
University of Wisconsin-Madison

Abstract

Why are some words more frequent than others? Some reasons are self-evident. A word like ”eat” is far more communicatively useful than a word like ”diagonalize”. But robust differences in frequency are also observed for words with seemingly equal communicative usefulness. For example, *hot* and *cold* seem equally important for communicating temperature yet *hot* in English is more frequent than *cold*. We focused on antonym pairs such as these and sought to predict differences in frequency from the connection patterns of these words in a semantic network while controlling for predictors like the number of word senses. Two network properties predicted word frequency especially well: the number of connections the word and its surrounding words have, and the ability of the word to connect less interconnected words. These two network properties not only predicted present word frequency, but also predicted future frequency changes suggesting a potential causality relationship between network properties and word frequency. Overall, this study offers new insights into the underlying causes of differences in word frequency and highlights the importance of considering a network perspective when examining how word frequency evolves.

Keywords: word frequency; word association network; language evolution

Introduction

Words are the basic units of language. They allow us to represent and communicate concepts, ideas, and information to ourselves and others, and are an essential part of how we program and structure our thoughts (Lupyan & Bergen, 2016). The frequency with which a word occurs in language has been used as a key predictor in psycholinguistics for many decades (Broadbent, 1967; Gorman, 1961; Hall, 1954; Sumby, 1963), and for good reason. Frequency matters. It predicts how well people recognize (Brysbaert, Stevens, Mandera, & Keuleers, 2016; Brysbaert & New, 2009; Ferrand et al., 2010; Keuleers, Diependaele, & Brysbaert, 2010; Keuleers, Lacey, Rastle, & Brysbaert, 2012; Yap & Balota, 2009) and remember them (Arndt & Reder, 2002; Clark, 1992; Gregg, 1976; Meier, Rey-Mermet, Rothen, & Graf, 2013; Yonelinas, 2002), and when children learn them (Braginsky, Yurovsky, Marchman, & Frank, 2019). Words that are more frequent also tend to be the source of metaphor and semantic change and predicts word frequency is also more , metaphor aptness (Littlemore, Sobrino, Houghton, Shi, & Winter, 2018) and semantic extensions of existing forms to new meanings (Harmon & Kapatsinski, 2017; Winter & Srinivasan, 2022).

Interestingly, despite the extensive literature on how word frequency predicts various aspects of human cognition, there has been little effort to understand why some words are more frequent than others (Calude & Pagel, 2014, 2011). We investigate this question of why some words are more frequent than others not only because it's a genuinely fascinating scientific question, but also because it provides insight into the values, beliefs, biases, and attitudes of the people who use them. For example, the frequency of certain words may reflect the prevalence of certain topics or issues, as well as people's bias toward these topics/issues in a particular society (e.g., Charlesworth, Caliskan, & Banaji, 2022; Lewis & Lupyan, 2020; Wu & Dunning, 2018). Additionally, studying why words become frequent can help to identify cultural trends (Grieve, Nini, & Guo, 2017).

The question of why some words are more frequent than others has several plausible answers, but as we will see, none of these suffice. Perhaps most obviously, certain words express more important ideas and so are used more frequently. For example, the word “lamp” is less frequent than the word “water” because we can live without a lamp, but not without water. A moment's thought reveals many counterexamples: for most readers mammals are more important than birds. Yet the word ”bird” is far more frequent than the word ”mammal”–indeed it is precisely because mammals are so important that we tend to refer to them with more basic terms like ”dog”. So in this case, importance predicts more narrow semantic extension and lower frequency.

Another answer is that words used to express familiar and common concepts or ideas are often used more frequently than words that are used to express more specialized or niche concepts. For example, the word “love” is used much more frequently than the word “sonder”, which refers to the realization that each passerby is living a life as complex as one's own. However, this explanation is somewhat circular because the frequency of a word can influence the subjective estimates of conceptual familiarity (Noble, 1954). Another factor is ecological frequency (Regier, Carstensen, & Kemp, 2016). We may talk about some things more just because they are more common in our environment. But word frequencies often depart wildly from ecological frequencies. While ”red” is far more frequent than ”green” or ”yellow”, there are not actually more red things than green or yellow things in the world. It's true that ”red” may be used more because it's more attention-grabbing (Ladle, Jepson, Correia, & Malhado, 2019; Winter, Perlman, & Majid, 2018) or has higher rele-

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

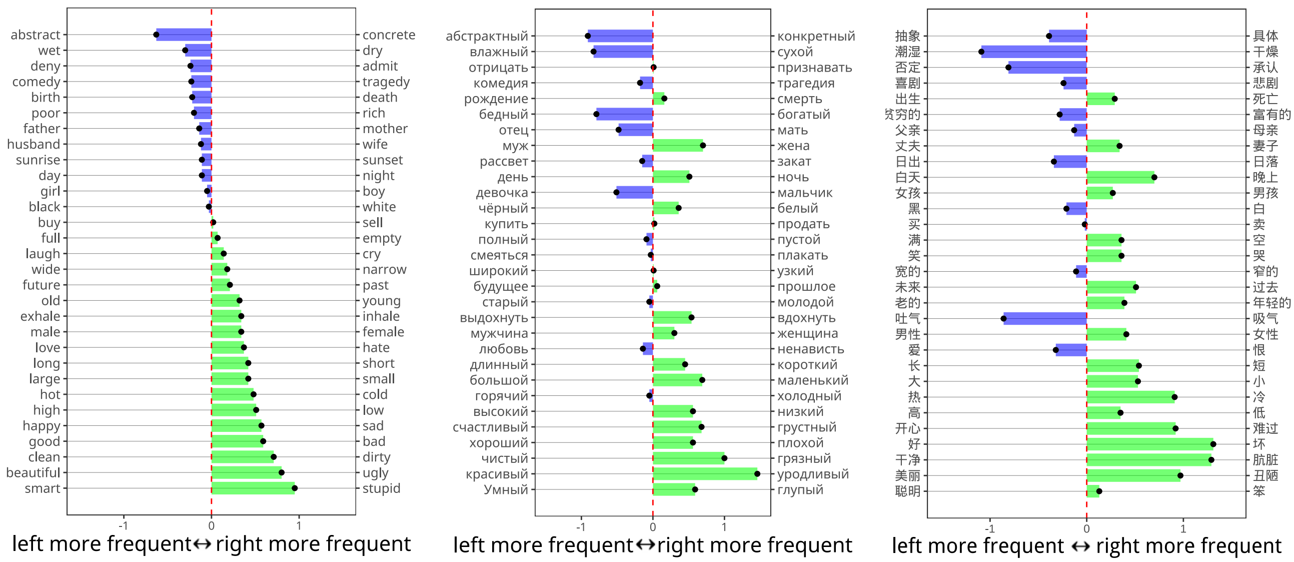


Figure 1: Thirty pairs of English antonyms and their difference in Zipf frequency in English (left), their Russian translation (middle), and their Mandarin Chinese translation (right). The cross-linguistic frequency data is obtained from Exquisite Corpus using the Python package *wordfreq*(Speer, 2022)

vance for survival (Blust, 2005), but such specific explanations are narrow, explaining only specific semantic domains.

Without discounting the importance of familiarity, salience, and communicative utility, we consider here a different way of investigating the causes of frequency by examining the place of different words in semantic networks. Past work has shown this approach to be useful in explaining various linguistic phenomena, most notably age of acquisition (Castro, Pelczarski, & Vitevitch, 2017; Hills, Maouene, Maouene, Sheya, & Smith, 2009; Kenett & Hills, 2022; Siew, Wulff, Beckage, & Kenett, 2019; Steyvers & Tenenbaum, 2005; Vitevitch, 2008), but not word frequency itself. For example, in a word association network, do words with more "central" positions tend to be more frequent? Does a certain position predict a word becoming more frequent over time?

Word association networks reflect the mental organization of words in human memory, based on the words' association strengths. The increased usage of a word might be attributed to the specific structural and functional roles it fulfils within this network, such as connecting to a broader range of contexts or bridging gaps between less connected words. So here, we explore why a word is more frequent by measuring different ways in which a word can be central in a semantic network. To help rule out alternative reasons for frequency such as psychological salience and communicative need, we focus on pairs of antonyms (i.e., words that are opposite in meaning) such as *male/female*, *full/empty*, *up/down* which are theoretically equivalent in terms of their function and other attributes yet almost never have the same frequency. Particularly, these antonyms varied in their frequency across different languages (e.g., as shown in Figure 1, thirty pairs of antonyms have correlated, but nevertheless diverging pattern of Zipf frequency (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014) in English, Russian, and Chinese), suggesting there is not a universal "ground truth" that causes one polarity to be more frequent than the other (we address the issue of word markedness below).

In the following section, we examined how differences in network centralities for pairs of antonyms predict differences in their frequency, both cross-sectionally and longitudinally. The hypotheses are: first, antonyms' differences in word frequency are predicted by their difference in network centralities during the same time period. Second, differences in network centralities at an earlier time predict meaningful changes in word frequency at a later time.

Cross-sectionally, do differences in network centralities predict differences in word frequency between antonyms?

We selected 774 antonym pairs of English words (see below) and modeled their difference in word frequency from differences in their network centralities.

Materials

We used three sources of data in this analysis. First, we used the word association network from the Small World of Words (SWOW) project (De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019), collected in the 2010s. This is a mega-corpus composed of crowdsourced word-association responses. Participants are shown target words and asked to respond with the first three words that come to mind. These responses were then given as cues to other participants to trigger further associates. We used this procedure to construct a weighted network with directed edges. The direction of edges indicates forward/backward associations and the weight is the probability of each backward/forward association given the response/cue. To focus on the dominant

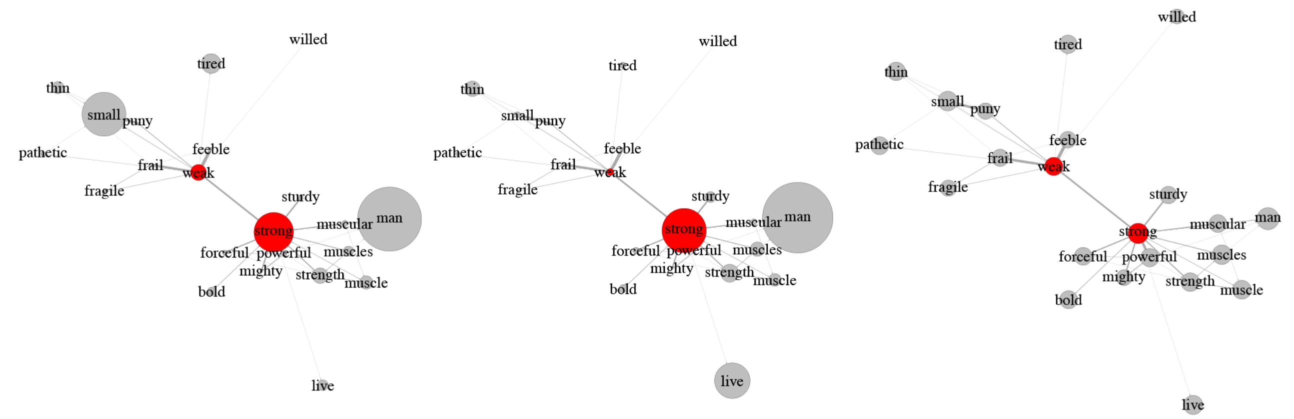


Figure 2: First-order associates of "strong" and "weak" in SWOW network, with the size of the node proportional to their degree centrality (left), betweenness centrality (middle), and closeness centrality (right).

translations, we used only the first response and excluded responses provided by only a single respondent. The network has 61935 words and 338839 directional associative links. Second, 3261 pairs of antonyms were originally extracted from WordNet (Fellbaum, 1998). We excluded those pairs that do not exist in SWOW network as well as those pairs that are not valid antonyms (e.g., *first* and *second*), and identified 774 pairs of antonyms (140 noun pairs, e.g., *male/female*, 457 adjectives pairs, e.g., *small/large*, and 177 verb pairs, e.g., *inhale/exhale*). Finally, Zipf frequency for each word was extracted from the Corpus of Historical American English (COHA) (Davies, 2022) during the decade of 2010s.

Variables

The dependent variable is the difference in Zipf frequency between 774 pairs of antonyms. The independent variables constitute differences in network centralities that measure how important a word is in a word association network.

There are several types of measures of a node's "importance": degree-based, distance-based, and neighborhood-based. The degree of a node refers to the number of edges connected to that node. For directed graphs, the degree can be further described as in-degree (i.e., the number of incoming edges), out-degree (i.e., the number of outgoing edges), and the sum of both in-degree and out-degree. Degree-based centrality generally considers words that have more direct connections as more central (e.g., degree-centrality: how many direct associates a word has), with some variations of degree-based centrality also taking into consideration how many neighbors those neighbors have (e.g., lobby index: the largest integer k such that a word has at least k associates with a degree of at least k).

Distance-based centrality considers words with short paths to others in the network as more central (e.g., closeness centrality, the average length of the shortest paths from a word to all other words). Neighborhood-based centrality considered words with more influence on its neighborhoods as more central. As opposed to distance-based centralities, which prioritize words that themselves have quick access to any other words, neighborhood-based centrality would consider words that function as bridge between otherwise less connected words as important. For example, a word can be important by having a lot of shortest paths going through it (i.e., betweenness centrality) or being involved in different cliques (i.e., cross-clique connectivity). Likewise, a word can also be important by having less mutually strongly connected neighbors so that they are less constrained by their neighbors (i.e., Burt's Constraint). Figure 2 shows a subnetwork on first-order associates of *weak* and *strong* (extracted from SWOW network). The size of each node is plotted in proportion to its degree centrality, betweenness centrality, and closeness centrality respectively. For example, *strong* (Zipf frequency = 5.22) has a higher word frequency than *weak* (Zipf frequency = 4.66), and it correspondingly has more connections to other words, being more important to its neighbors such that the shortest paths between other words will largely increase if strong doesn't exist in the network, and is slightly closer to other words.

Our independent variables are 24 scores computed from 12 network attributes between each pair of antonyms. The 12 network attributes are either degree-based centralities (e.g., degree centrality, coreness, diffusion degree, lobby index, Laplacian centrality, alpha centrality, and PageRank centrality), distance-based centralities (e.g., closeness centrality, radiality centrality), or neighborhood-based centralities (e.g., betweenness centrality, Burt's constraint, cross-clique connectivity). The measures of centrality we investigated here are common ones, but do not exhaust ways of measuring centrality. Figure 3 shows that centralities from the same class or based on the same direction of association (i.e., backward/forward) are highly correlated.

In addition to network centralities as main parameters of interest, we controlled for three covariates: (1) The difference in the number of morphemes (derived words such as *unclear* may be less frequent than the base form *clear* because it is more complex); (2) Words with more senses have obviously more opportunities for use which drives up word frequency (e.g., *fall* means both "move downward" which is opposed to *rise*, but also means "autumn"). (3) The difference

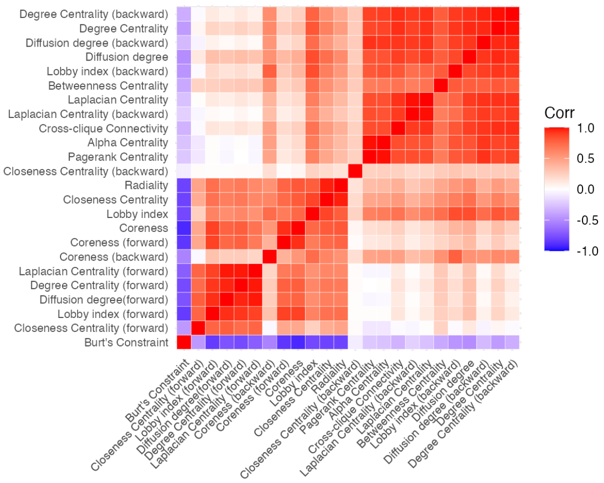


Figure 3: Correlation between network centralities from SWOW networks based on selected antonyms.

in frequency of a word being mentioned as a response to any cue (i.e., the words frequency in SWOW). The main reason to control for the SWOW frequency is that a more frequent word also tends to be mentioned more as a response, and degree-based measures are partially reflecting this frequency effect. Controlling for SWOW frequency allows us to test the effect of network centralities above and beyond the frequency effect that centrality measures may inevitably capture and avoid to make circular inference using frequency to predict frequency.

Results

Given the multicollinearity between network centralities (see Figure 3), we regress the difference in word frequency on each network centrality one at a time while controlling for differences in number of morphemes, differences in number of senses, and differences in SWOW frequency. Noted that many centralities were highly skewed (e.g., degree (+backward), lobby index (+backward), betweenness, PageRank, Burt's constraint, alpha, Laplacian (+backward), diffusion, diffusion (+backward), coreness (+backward), cross-clique connectivity), so we log-transformed them to correct for the high skewness of their distribution.

As shown in Figure 4(a), most differences in network centralities are significant predictors of differences in word frequency. Except alpha, PageRank, and coreness (backward), all degree-based centralities are consistently positive predictors of word frequency ($p < .001$), indicating that compared to its antonyms, the more (forward/backward) associations a word and its neighbors have, the more frequent the word is. Meanwhile, distance-based centralities (except Radiality) are positive predictors of word frequency as well ($p < .05$), suggesting that compared to its antonyms, words with shorter distances to other words are more frequent. Neighborhood-based network centralities are also significant predictors of word frequency ($p < .01$). Specifically, positive coefficients for betweenness centrality and cross-clique connectivity, and the negative coefficient for Burt's constraint (indicates that words with fewer strongly interconnected, redundant neighbors have higher frequencies) all suggest words playing a role of "bridge" in word association network tend to be more frequent.

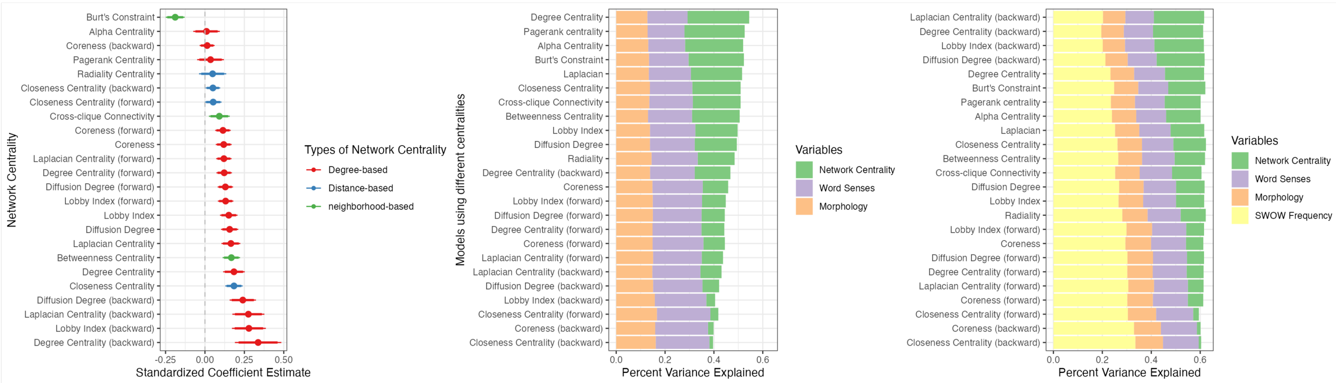
People are more likely to name frequent words so a word with higher degree-based centralities(e.g., have more backward associations) tends to be more frequent. Given that frequency is an essential part of setting up the network structure, it's surprising to find that network centralities may account for substantial amount of variance even after controlling for SWOW frequency (how frequent a word is given as a response to any cue). As shown in Figure 4 (b)-(c), adding SWOW frequency as a covariate admittedly reduced the variance explained by network centralities, but the network centralities can explain frequency above and beyond the frequency effect inevitably embedded in them.

The results of the analysis above indicate that network centralities can significantly predict word frequency data collected at a similar point in time. However, it should be noted that this analysis does not allow for the determination of causality. It is not clear whether word frequency is the cause or result of a word's network properties, and whether network properties at an earlier time can predict subsequent changes in word frequency.

Using network centrality to predict how word frequencies change in the future

To examine whether network properties drive changes in word frequency, we investigate whether differences in network centralities at an earlier time predict later differences in frequency change between the antonyms.

The methods are nearly identical to those used in the previous section, with three exceptions. First, the SWOW word association network is substituted by the University of South Florida Free Association (USF) Norms (Nelson, McEvoy, & Schreiber, 2004), collected from 1970s to 1990s. In this study, a single-response procedure was used in which participants were asked to respond with the first word that came to mind. Again, we excluded all idiosyncratic responses. The USF network has 10616 words and 72162 directional associative links. Since it's much smaller (about 1/6 the size of the SWOW network), 520 pairs of antonyms were found in this network (96 nouns, 293 adjectives, and 131 verbs). Second, the dependent variables are no longer differences in word frequency between antonyms. Rather, the dependent variable is the differences in word frequency change from an earlier time 1970s, approximately from which the USF norms were originally collected) to a later time (2010s) between each pair of antonyms. For example, the Zipf frequency of *good* changes from 6.092 (1970s) to 6.039 (2010s), and its antonym *bad* changes from 5.416 to 5.465. The dependent variable for this good/bad would be $(6.092 - 6.039) - (5.416 - 5.465) = 0.1$. Third, instead of controlling for the difference in SWOW frequency, we controlled for the difference in word frequency as



(a) (b) (c)

Figure 4: (a). Difference in antonyms' network centralities (SWOW) predicts their difference in COHA word Zipf frequency data obtained from 2010s, controlled for difference in senses, morphemes, and SWOW frequency. (b). Percent of variance explained by variables for models, including differences in senses, morphology, and network centralities. (c) Percent of variance explained by variables for models, including differences in senses, morphemes, network centralities, and SWOW frequency.

estimated the 1970s COHA corpus to test the effect of network centrality beyond frequent words becoming even more frequent.

Analysis & result

Word frequencies for 520 pairs of antonyms were very stable. The average intercorrelation between word frequency from 1970s, 1980s, 1990s, 2000s, 2010s was as high as r = .99. We regressed the small difference in frequency change on network centralities (again, with the highly skewed ones log-transformed), controlled for differences in number of morphemes, differences in number of senses, and differences in word frequency from the 1970s. The results are shown in Figure 5.

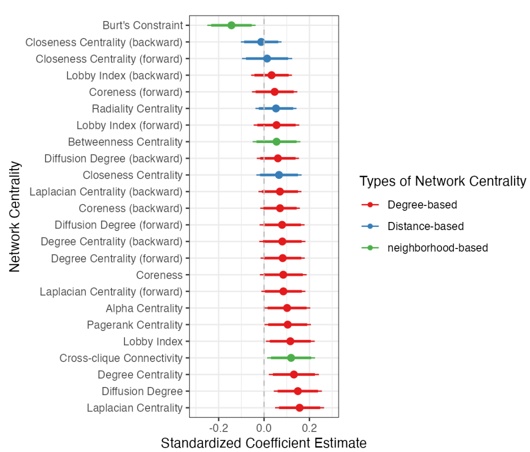


Figure 5: Difference in antonyms' network centralities (USF) predict their difference in COHA word Zipf frequency change from 1970s to 2010s

For degree-based centralities, differences in laplacian centrality ($p < .005$), diffusion degree ($p < .01$), degree centrality ($p < .05$), lobby index ($p < .05$), pagerank centrality ($p < .05$) from the USF network are significant predictors of difference in frequency change until the 2010s. Differences in alpha centrality, Laplacian centrality (forward), coreness, degree centrality (forward) are marginally significant ($p <= .1$) Though the effect sizes for significant (or marginally significant) degree-based centralities are small (which is expected given the minute fluctuation in word frequencies for these words in the studied time period and a much smaller network size compared to SWOW), the directions of prediction are all as expected: compared to its antonym, the more connections a word has at an earlier time point, the more frequent it will grow to be in the following decades. Differences in neighborhood-based centralities such as Burt's Constraint and cross-clique connectivity ($p < .001$) are significant predictors of longitudinal word frequency change, and the interpretation of its negative coefficient is consistent with the cross-sectional analysis: The words that fill structural holes or bridge otherwise less connected words tend to become more frequent in the future. Distance-based centralities such as closeness and radiality are not significant predictors for differences in frequency change ($p > .2$), indicating they are more likely to be solely the consequence rather than the cause of frequency change. Alternatively, distance-based centralities like closeness centrality is a noisier measure since they are more sensitive to outliers.

Discussion

The question of why some words are more frequent than others does not have a single answer and includes ecological and social relevance as well as psychological prominence–all difficult-to-quantify measures. We examined whether properties of semantic networks (created using word associations) can (1) predict which words of roughly similar prominence and relevance are more frequent, and (2) whether these measures are mere consequences of word frequencies or help predict how they change – consistent with a causal effect. Our results support two general conclusions: First, the more connections a word and its neighbors have, the higher its frequency. This could mean that the more frequent the word, the higher the chance it has to be associated with other words, leading to higher degree-based centralities (e.g., degree centrality, lobby index). Alternatively, the more connections a word and its neighbors has (hence higher degree-based centrality), the higher its frequency because it has more chances to be used. The longitudinal analysis suggests that greater centrality at time 1 words predicts a higher frequency at time 2, lending some support to the second interpretation. A supplementary analysis actually suggested differences in word frequency between antonyms from the 1970s can predict changes in degree-based centrality when comparing earlier USF norms to later SWOW norms collected decades apart[1]. Altogether, both interpretations may be true: the more frequent the word, the higher its degree-based centralities, and the word becomes even more frequent by benefiting from its central positions—a phenomenon where the "the rich get richer", also known as frequency-dependent selection and the Matthew effect (Merton, 1968).

The second conclusion is that words with higher influence on their neighbors (e.g., those mediating the shortest paths between other words or connect different cliques) have higher frequencies and will likely have even higher frequencies in the future. In social network studies an agent in such position is said to fill a "structure hole". It's shown that agents that fill structure holes are more likely to be provided with more social capital such as good, creative ideas ((Burt, 2004). In the case of words as agents, one implication of this finding is that words that fill "structural hole" positions in an associative network may benefit from the semantic diversity of their neighbors. These neighbors might, for example, expand the word's semantic coverage. This finding also aligns with the observation that words with wider topical dissemination become more entrenched in the lexicon, while those limited to narrower contexts are more prone to falling out of use (Altmann, Pierrehumbert, & Motter, 2011; Stewart & Eisenstein, 2017; Francis, Rabinovich, Samir, Mortensen, & Stevenson, 2021).

An alternative explanation of our findings is that the more frequent words in each antonym pair correspond to a default or unmarked end of the dimension (Clark, 1992). For example, "tall" is treated as the unmarked end of tall-short. Asking "how tall is he?" does not imply tallness while asking "how short is he?" has the implication of shortness. The unmarked end is typically more frequent and more easily processed (Proctor & Cho, 2006). But *why* does "tall" get to be unmarked? Recall also that there are cross-linguistic differences in which word in the antonym pair is more frequent. Might markedness itself be a function of a word's position in the semantic network?

Our work adds to the current understanding of the factors influencing the frequency of words, suggesting that a word might become more frequent just by virtue of being well-connected or in a position of a structural hole in a word association network. Past research shows that language evolves to satisfy communication needs (Regier et al., 2016). Factors such as importance, familiarity, salience, and utility align well with this motivation, as they explained why words being more "useful" for efficient communication might be promoted as more frequent. It's less clear whether the word association dynamic is unidirectionally driven by communicative needs, or–more interestingly–if a word's position can change communicative need, helping to create communicative situations where the word comes in handy.

Another unanswered question is what gives rise to a word's network centrality? We believe that centrality is a function of both salient verbal and non-verbal experiences, which are reflected in word association norms . For instance, the cue word *happy* could elicit the response *smile*, as both *smile* cooccur in language, but could also reflect reliance on imagery, emotion and other non-verbal experiences (see for instance, De Deyne et al., 2021). However, the specific factors that contribute to a word's network centralities still require further investigation.

Finally, we focused here on pairs of antonyms to help rule out the possibility for the network effects to be confounded by well-known contributors to word frequency, e.g., basic-level terms being more frequent than super- and subordinate terms, words naming more frequently encountered things ("dog") being more frequent than words denoting less frequently encountered things ("wombat"). Future investigations need to understand whether the predictors identified here apply to a wider range of words and how network centralities compare to more conventional predictors such as familiarity, and communicative utility (these latter measures are, however, difficult to quantify). We also predict that cross-linguistic differences in frequency should be predicted by cross-linguistic differences in the words' position in semantic networks–a prediction whose testing requires building analogous semantic networks in multiple languages.

Conclusion

We predicted the differences in word frequency between pairs of antonyms from the difference in 12 network centrality measures cross-sectionally and longitudinally. We show that the more connections a word and its neighbors have, the more frequent the word is, and will be. Words playing a bridging role and those filling a structural hole in word associations also tend to be more frequent now and arepredicted to become more frequent in the future. Our study is the first attempt to understand word frequency from the dynamics of a word association network and contributes an additional element to the current understanding to the origins and evolution of word frequencies.

References

Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2011). Niche as a determinant of word fate in online groups. *PloS one*, *6*(5), e19009.

[1] We did not include this analysis because the SWOW and USF datasets were collected using different procedures and reflect different populations meaning that the SWOW network cannot be treated as simply a later version of the USF network

Arndt, J., & Reder, L. M. (2002, September). Word Frequency and Receiver Operating Characteristic Curves in Recognition Memory: Evidence for a Dual-Process Interpretation. *Journal of experimental psychology. Learning, memory, and cognition*, *28*(5), 830–842. doi: 10.1037//0278-7393.28.5.830

Blust, R. (2005). Liver and Lungs: A Semantic Dyad in Austronesian Languages. *Oceanic Linguistics*, *44*(2), 537–543. (Publisher: University of Hawai'i Press)

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019, June). Consistency and Variability in Children's Word Learning Across Languages. *Open Mind: Discoveries in Cognitive Science*, *3*, 52–67. doi: 10.1162/opmi$_{a0}$0026

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological review*, *74*(1), 1.

Brysbaert, M., & New, B. (2009, November). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. doi: 10.3758/BRM.41.4.977

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 441–458. (Place: US Publisher: American Psychological Association) doi: 10.1037/xhp0000159

Burt, R. (2004, September). Structural Holes and Good Ideas. *American Journal of Sociology*, *110*(2), 349–399. (Publisher: The University of Chicago Press) doi: 10.1086/421787

Calude, A. S., & Pagel, M. (2011, April). How do we use language? Shared patterns in the frequency of word use across 17 world languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1101–1107. Retrieved 2023-01-31, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049087/` doi: 10.1098/rstb.2010.0315

Calude, A. S., & Pagel, M. (2014). *Frequency of use and basic vocabulary* (L. Filipović & M. Pütz, Eds.). John Benjamins Publishing Company. Retrieved 2023-01-31, from `https://benjamins.com/catalog/hcp.44.02cal` doi: 10.1075/hcp.44.02cal

Castro, N., Pelczarski, K. M., & Vitevitch, M. S. (2017, July). Using Network Science Measures to Predict the Lexical Decision Performance of Adults Who Stutter. *Journal of Speech, Language, and Hearing Research*, *60*(7), 1911–1918. doi: 10.1044/2017$_{JSLHR-S-16-0298}$

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. In *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510–516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Charlesworth, T. E., Caliskan, A., & Banaji, M. R. (2022). Historical representations of social groups across 200 years of word embeddings from google books. *Proceedings of the National Academy of Sciences*, *119*(28), e2121798119.

Clark, S. E. (1992, May). Word frequency effects in associative and item recognition. *Memory & Cognition*, *20*(3), 231–243. doi: 10.3758/BF03199660

Davies, M. (2022). *Corpus of Historical American English (COHA).* Abacus Data Network. doi: 11272.1/AB2/EQQKSQ

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "small world of words" english word association norms for over 12,000 cue words. *Behavior research methods*, *51*, 987–1006.

Feigenbaum, E. A. (1963). The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill.

Fellbaum, C. (1998). *WordNet: An electronic lexical database and some of its applications.* MIT press Cambridge.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., … Pallier, C. (2010, May). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, *42*(2), 488–496. doi: 10.3758/BRM.42.2.488

Francis, D., Rabinovich, E., Samir, F., Mortensen, D., & Stevenson, S. (2021). Quantifying cognitive factors in lexical decline. *Transactions of the Association for Computational Linguistics*, *9*, 1529–1545.

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of experimental psychology*, *61*(1), 23.

Gregg, V. (1976). Word frequency, recognition and recall. In *Recall and recognition* (pp. x, 275–x, 275). Oxford, England: John Wiley & Sons. doi: 10.1037/11314-016

Grieve, J., Nini, A., & Guo, D. (2017, March). Analyzing lexical emergence in Modern American English online1. *English Language & Linguistics*, *21*(1), 99–127. Retrieved 2023-02-02, from `https://www.cambridge.org/core/journals/english-language-` doi: 10.1017/S1360674316000113

Hall, J. F. (1954). Learning as a function of word-frequency. *The American journal of psychology*, *67*(1), 138–140.

Harmon, Z., & Kapatsinski, V. (2017, November). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22–44. doi: 10.1016/j.cogpsych.2017.08.002

Hill, J. A. C. (1983). A computational model of language acquisition in the two-year old. *Cognition and Brain Theory*, *6*, 287–317.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009, June). Longitudinal Analysis of Early Semantic Networks: Preferential Attachment or Preferential Acquisition? *Psychological Science*, *20*(6), 729–739. (Publisher: SAGE Publications Inc) doi: 10.1111/j.1467-9280.2009.02365.x

Kenett, Y. N., & Hills, T. T. (2022). Editors' Introduction to Networks of the Mind: How Can Network Science Elucidate Our Understanding of Cognition? *Topics in Cognitive Science*, *14*(1), 45–53. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tops.12598) doi: 10.1111/tops.12598

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: a lexical decision study on 14,000 dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, *1*, 174. doi: 10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012, March). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304. doi: 10.3758/s13428-011-0118-4

Ladle, R. J., Jepson, P., Correia, R. A., & Malhado, A. C. M. (2019). A culturomics approach to quantifying the salience of species on the global internet. *People and Nature*, *1*(4), 524–532. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pan3.10053) doi: 10.1002/pan3.10053

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

Littlemore, J., Sobrino, P. P., Houghton, D., Shi, J., & Winter, B. (2018, April). What makes a good metaphor? A cross-cultural study of computer-generated metaphor appreciation. *Metaphor and Symbol*, *33*(2), 101–122. (Publisher: Routledge _eprint: https://doi.org/10.1080/10926488.2018.1434944) doi: 10.1080/10926488.2018.1434944

Lupyan, G., & Bergen, B. (2016). How Language Programs the Mind. *Topics in Cognitive Science*, *8*(2), 408–424. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tops.12155) doi: 10.1111/tops.12155

Matlock, T. (2001). *How real is fictive motion?* Doctoral dissertation, Psychology Department, University of California, Santa Cruz.

Meier, B., Rey-Mermet, A., Rothen, N., & Graf, P. (2013). Recognition memory across the lifespan: the impact of word frequency and study-test interval on estimates of familiarity and recollection. *Frontiers in Psychology*, *4*.

Merton, R. K. (1968). The matthew effect in science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004, August). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. doi: 10.3758/BF03195588

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Noble, C. E. (1954). The familiarity-frequency relationship. *Journal of Experimental Psychology*, *47*, 13–16. (Place: US Publisher: American Psychological Association) doi: 10.1037/h0060025

Ohlsson, S., & Langley, P. (1985). *Identifying solution paths in cognitive diagnosis* (Tech. Rep. No. CMU-RI-TR-85-2). Pittsburgh, PA: Carnegie Mellon University, The Robotics Institute.

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

Regier, T., Carstensen, A., & Kemp, C. (2016, April). Languages Support Efficient Communication about the Environment: Words for Snow Revisited. *PLOS ONE*, *11*(4), e0151138. (Publisher: Public Library of Science) doi: 10.1371/journal.pone.0151138

Shrager, J., & Langley, P. (Eds.). (1990). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.

Siew, C. S. Q., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019, June). Cognitive Network Science: A Review of Research on Cognition through the Lens of Network Representations, Processes, and Dynamics. *Complexity*, *2019*, e2108423. (Publisher: Hindawi) doi: 10.1155/2019/2108423

Speer, R. (2022, September). *rspeer/wordfreq: v3.0.* Zenodo. doi: 10.5281/zenodo.7199437

Stewart, I., & Eisenstein, J. (2017). Making" fetch" happen: The influence of social and linguistic context on nonstandard word growth and decline. *arXiv preprint arXiv:1709.00345*.

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, *29*(1), 41–78. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog2901_3) doi: 10.1207/s15516709cog2901$_3$

Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Memory and Language*, *1*(6), 443.

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-uk: A new and improved word frequency database for british english. *Quarterly journal of experimental psychology*, *67*(6), 1176–1190.

Vitevitch, M. S. (2008, April). What Can Graph Theory Tell Us About Word Learning and Lexical Retrieval? *Journal of Speech, Language, and Hearing Research*, *51*(2), 408–422. (Publisher: American Speech-Language-Hearing Association) doi: 10.1044/1092-4388(2008/030)

Winter, B., Perlman, M., & Majid, A. (2018, October). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, *179*, 213–220. doi: 10.1016/j.cognition.2018.05.008

Winter, B., & Srinivasan, M. (2022, January). Why is Semantic Change Asymmetric? The Role of Concreteness and Word Frequency and Metaphor and Metonymy. *Metaphor and Symbol*, *37*(1), 39–54. (Publisher: Routledge _eprint: https://doi.org/10.1080/10926488.2021.1945419) doi: 10.1080/10926488.2021.1945419

Wu, K., & Dunning, D. (2018, March). Hypocognition: Making Sense of the Landscape beyond One's Conceptual Reach. *Review of General Psychology*, *22*(1), 25–35. Retrieved 2023-02-02, from `https://doi.org/10.1037/gpr0000126` doi: 10.1037/gpr0000126

Yap, M. J., & Balota, D. A. (2009, May). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502–529. doi: 10.1016/j.jml.2009.02.001

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517. (Place: Netherlands Publisher: Elsevier Science) doi: 10.1006/jmla.2002.2864