Evolution on the Lexical Workbench: Disentangling Frequency, Centrality, and Polysemy in Language Evolution

Zach Studdiford† **(studdiford@wisc.edu)**
Department of Psychology,
University of Wisconsin–Madison
Department of Computer Science,
University of Wisconsin–Madison

Qiawen Liu† **(ella.qiawen.liu@princeton.edu)**
Department of Psychology,
University of Wisconsin–Madison
Department of Computer Science,
Princeton University

Ying Li (liying@psych.ac.cn)
Institute of Psychology,
Chinese Academy of Sciences
State Key Lab of Cognitive Science

Zhengjun Zhang (zjz@stat.wisc.edu)
Department of Statistics,
University of Wisconsin–Madison

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

Abstract

How do words evolve in their usage and meaning over time? We investigate the relationship between word frequency, semantic richness, and network centrality through longitudinal analysis of the Corpus of Historical American English (1820-2019). Using measures of semantic richness and network position, we find that a word's betweenness centrality—its tendency to bridge different semantic domains—consistently predicts both its future semantic richness and frequency of use. This relationship strengthens over longer time intervals, with the strongest effects observed across a 100-year span. Notably, while frequency and semantic richness are correlated as established in the literature, our results indicate that there was no directional relationship between frequency and semantic richness, while network centrality exerts a significant influence on both of these factors. Our results suggest that a word's position within the semantic network might play a crucial role in its evolution: words that bridge different semantic domains are more likely to develop new meanings and change in frequency over time. These findings offer new insights into the mechanisms driving language change.

Keywords: language evolution; frequency; network; semantic extension

Introduction

"A struggle for life is constantly going on among the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their own inherent virtue." (Darwin (1872), quoting Müller (1870) referencing Schleicher (circa 1860))

Scholars have long invoked evolutionary metaphors to capture how languages change over time. Indeed, language is constantly evolving, shaped by the collective choices of thousands of speakers across generations. Open a book from centuries ago, and you'll find some words, like "wherefore" and "thou," have virtually vanished from modern speech, while others, like "mother" and "water," remain as vital as ever. Words like "meat" (once meaning any kind of food) narrowed in meaning, while "dog" (originally referring to a specific breed) broadened to refer to all dogs. The meaning of "awful" shifted from "awe-inspiring" to terrible while "nice" shifted from foolish to, well, nice. These shifts are a natural part of language evolution, and much like species in nature, words rise and fall, expand or shrink, persist or fade as they adapt to serve our ever-shifting communication needs (Gibson et al., 2019; Mahowald et al., 2018; Trott & Bergen, 2022).

How do words' usage and meanings evolve over time? Several interrelated characteristics play a role in this process, with word frequency and polysemy standing out as the most studied factors (Crossley et al., 2010; Piantadosi et al., 2012; Zipf, 1945). Polysemy refers to the phenomenon where a word acquires multiple meanings (Cruse, 1986). Research suggests that as words become more frequent in speech, they tend to develop additional meanings, expanding their range of usage (Harmon & Kapatsinski, 2017). Zipf's tool metaphor captures this dynamic well: he likened linguistic forms to tools used by speakers to accomplish communicative tasks (Zipf, 1949). Just as an artisan selects tools based on their accessibility and versatility, speakers use words that are easily accessible and can serve many functions. The more frequently a word is used, the more meanings it tends to acquire, making it increasingly adaptable to new contexts and communication needs. As a result, words that are used most often are often the most polysemous, acquiring meanings to meet diverse communicative demands (Bybee, 2010; Harmon & Kapatsinski, 2017; Piantadosi et al., 2012).

Yet focusing on the functionality of individual words captures only part of the picture. Just as species evolve not solely in response to the physical environment but also in relation to one another (Ehrlich & Raven, 1964; Thompson, 1994), words may evolve through their connections within a larger semantic ecosystem. While frequency reflects how often a word is used, *centrality* describes how well-connected it is to the broader network of meanings. Words that occupy central positions—akin to keystone species—are more likely to be encountered and thus may evolve in both frequency *and* meaning (Liu et al., 2024; Liu et al., 2023; Steyvers & Tenenbaum, 2005). In Zipf's tool metaphor, this is like placing a favorite tool in the most convenient spot on the workbench: it becomes the easy, go-to choice for many tasks, reinforcing its utility. Similarly, a central word is primed for use in diverse

† Joint first authors.

Full code and measures data available at: https://github.com/zstud04/lexical-evolution

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

contexts, further solidifying its frequency and paving the way for new meanings to emerge.

This interplay between frequency, polysemy, and network centrality raises foundational questions about their directional relationships. Are words more likely to become central as they are already frequent, or do they become frequent by virtue of their central positions in the network? Are more polysemous words naturally more central, or does occupying a central hub encourage multiple meanings to proliferate? Untangling these possibilities is crucial for understanding language evolution—specifically, whether these forces operate in a mutually reinforcing cycle or whether one exerts an asymmetric influence on the others.

To address this challenge, we propose a novel approach based on longitudinal corpus analysis. By analyzing the evolution of word usage over time, we aim to disentangle the complex interactions between frequency, network centrality, and polysemy. We introduce new methods for calculating centrality from corpus data, alongside measures of the richness of meaning in adjectives, to explore how these dimensions co-evolve. Through this approach, we aim to uncover how words' positions within a network and their flexibility in meaning go together with their rise and fall in language.

Approach

Our data comes from the Corpus of Historical American English (COHA), which spans 1820–2019 and includes various genres (e.g., fiction, newspapers, and magazines). By examining language usage in a wide temporal window, we can track how linguistic factors shift and interact across time.

Word Frequency

We begin by extracting raw frequencies of each word per decade from COHA. Raw frequencies typically follow a heavy-tailed (Zipfian) distribution: a small number of words occur with extremely high frequency, while the vast majority of words appear rarely. This distributional characteristic poses challenges for interpretation as well as statistical modeling, particularly for methods that assume linear relationships.

To address these challenges, we implement a two-step normalization process. First, we normalize the raw counts by the total number of tokens in each decade to obtain frequencies per million words (*pmw*). Then, following Van Heuven et al. (2014), we apply a logarithmic transformation to obtain values on the Zipf scale:

$$\mathrm{Zipf}(w) = \log_{10}\big(\mathrm{freq}_{\mathrm{pmw}}(w)\big) \tag{1}$$

where $\mathrm{freq}_{\mathrm{pmw}}(w)$ represents the frequency of word w per million words.

Semantic Richness

We operationalize semantic richness by evaluating the variety of contexts in which words (specifically adjectives) appear. Adjectives are used as the target for the analysis, as the semantic context of an adjective can be directly inferred from the noun it modifies in a sentence. Our analysis focuses primarily on the fiction subcorpus in COHA, as the creative language in fiction can reveal nuanced semantic shifts. For example, the adjective *small* can be used in physical descriptions (*small house*, *small town*) in addition to more abstract phrases (*small talk*), highlighting the kind of semantic variation we aim to capture.

We use `spaCy` (Honnibal & Montani, 2017) to extract all phrases where an adjective modifies a noun in each decade of the corpus, grouping by unique adjectives. For each adjective, we collect the set of isolated noun-adjective pairs (e.g., "small house", "small talk"). We then compute a *contextual embedding* for the adjective word within each pair, capturing its meaning in the context of the noun. These embeddings are assembled into a matrix, from which we compute four complementary measures.

1. **Spectral Diversity.** Spectral diversity refers to how *spread out* a word's meaning is across different semantic directions. More formally, let M be an $N \times d$ matrix, where N represents the number of occurrences and d the embedding dimensions. The singular values of M quantify the variance captured along each principal direction. A larger range of singular values (i.e., the difference between the largest and smallest singular value) indicates higher spectral diversity, reflecting wider semantic contexts.

2. **Number of Nonzero Eigenvalues.** The number of nonzero eigenvalues of $M^\top M$ corresponds to the *effective rank* of the matrix, reflecting how many distinct semantic dimensions (or clusters of usage) the adjective occupies. A higher count indicates a broader range of contexts.

3. **Entropy.** We also assess the distribution of contexts using an *entropy* measure based on the eigenvalues of $M^\top M$. Conceptually, this captures how evenly variance is distributed across different dimensions in the embedding space. If an adjective is used in many different senses with roughly equal frequency, the eigenvalue distribution is more uniform, leading to higher entropy. Conversely, if the variance is concentrated in one or two dimensions (indicating predominant usage in one or two senses), the entropy is lower.

4. **Condition Number.** The condition number of $M^\top M$ is the ratio of its largest to its smallest singular value. It captures whether the different semantic dimensions are balanced (lower condition number) or dominated by one dimension (higher condition number).

Adjectives with richer meanings will modify a more diverse set of nouns, resulting in embeddings that occupy a greater volume of the high dimensional semantic space. Thus, by comparing the embeddings of *small* in contexts like *small town*, *small house*, and *small talk*, we can observe how meanings cluster and diverge over time, thereby capturing the evo-

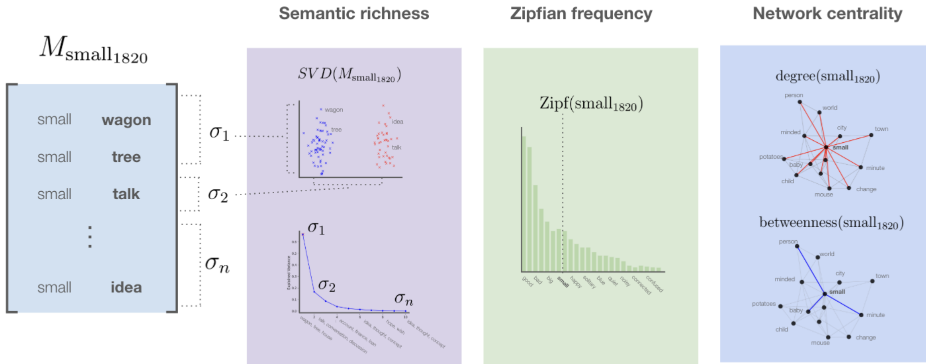

Figure 1: Example illustration of word properties for *small* in 1820. The principal axes are obtained through spectral value decomposition (SVD) of the embeddings matrix, which reflect the general semantic directions or "senses" of an adjective. Nouns are assigned to the semantic sense for which their embeddings vector has the highest projection. A plot of the explained variance for each principal axis shows the extent to which one semantic direction dominates the meaning of the adjective.

lution of semantic richness through these matrix-based measures. Figure 1 illustrates how different measures of semantic richness, word frequency, and network centrality are obtained for a given word from the adjective-noun matrix.

Network Centrality

In addition to frequency and semantic richness, we also examine how words' *positions* in a semantic network shift over time. To construct these networks:

1. **Co-Occurrence Extraction.** We define nodes as individual words and draw edges between words that co-occur in the corpus within a 5-token sliding window.

2. **PPMI Transformation.** Raw co-occurrence counts are converted into Positive Pointwise Mutual Information (PPMI) scores (Church & Hanks, 1990):

$$\mathrm{PPMI}(v_i, v_j) = \max\left(0, \log\left(\frac{p(v_i, v_j)}{p(v_i)\, p(v_j)}\right)\right),$$

   where $p(v_i, v_j)$ is the probability of words v_i and v_j co-occurring within the window, and $p(v_i)$, $p(v_j)$ are their individual empirical probabilities. The resulting PPMI matrix acts as a weighted adjacency matrix for the semantic network.

From the PPMI-weighted network, we compute 14 centrality measures highlighting complementary aspects of node importance. These include *betweenness centrality* (measuring a node's role in briding shortest paths between other nodes), *degree centrality* (a count of a node's immediate neighbors), and *dmnc* (the connectivity of a node's immediate neighbors).

Time-series analysis

To examine the relationships between word frequency, centrality, and semantic richness, we compile time-series data for each measure across all decade intervals from 1820 to 2000. To assess whether a measure X has an effect on a measure Y at time $t+k$ (where k is the lagged interval) , we apply three criteria used in causal inference (Pearl, 2009):

1. **Temporal asymmetry**: X_t predicts Y_{t+k} more than Y_t predicts X_{t+k}, and earlier X_t is a better predictor of Y_{t+k} than X_{t+k}.

2. **Granger causality**: X_t improves prediction of Y_{t+k} beyond past values of Y_t (Granger, 1969).

3. **Unique variance**: The conditional variance of Y_{t+k} is reduced when conditioning on X_t, isolating variability attributable to X_t and not variance or residual noise in Y_{t+k}. Critically, conditioning X_{t+k} on Y_t has no such effect.

The properties of our word measures data present several challenges for traditional time series methods such as Granger causality and vector autoregression. These include the high dimensionality of our data and a low number of observations (19 word properties over 12 decade instances), Non-stationarity in our time series, and heteroskedasticity in the residuals for measures of network centrality and semantic richness. To address this, we apply non-parametric *generalized measures of correlation* (GMC) to measure dependencies between word properties without assuming linearity or stationarity (Zheng et al., 2012). GMC measures the proportion of variance in one variable uniquely explained by conditioning on another, extending Pearson's correlation to capture asymmetric dependencies between variables. GMC builds on the standard variance decomposition formula:

$$\mathrm{Var}(Y) = \mathrm{E}\big[\mathrm{Var}(Y \mid X)\big] + \mathrm{Var}\big(\mathrm{E}[Y \mid X]\big),$$

which partitions the total variance of Y into the expected conditional variance (variance of Y not explained by X) and the variance of the conditional expectation (variance of Y explained by X). From this, the proportion of variance in Y explained by X can be expressed as:

$$1-\frac{\mathrm{E}\left[\mathrm{Var}(Y \mid X)\right]}{\mathrm{Var}(Y)},$$

quantifying how much conditioning on X reduces the total variance of Y. GMC then expresses this in terms of the expected squared deviation from the conditional mean:

$$\mathrm{GMC}(Y \mid X)=1-\frac{\mathrm{E}\left[(Y-\mathrm{E}[Y \mid X])^2\right]}{\mathrm{Var}(Y)},$$

where centering Y on $\mathrm{E}[Y \mid X]$ isolates the variance in Y that is explained by X. $\mathrm{E}[Y \mid X]$ is estimated non-parametrically using kernel smoothing to flexibly capture nonlinear dependencies. This framework satisfies our criteria for causal inference: GMC isolates unique variance in Y attributable to X, remains robust under nonlinearity and non-stationarity, and allows for a test of directionality by comparing $\mathrm{GMC}(Y \mid X)$ to $\mathrm{GMC}(X \mid Y)$.

Analysis

To test the influence of word frequency, semantic richness, and network centrality on each other over time, we applied GMC analysis to measures obtained for adjectives from COHA in decades 1820 to 2000. After identifying adjective-noun pairs in each corpus, we filtered for adjectives appearing in all decades and without NaN values for any matrix calculations. This resulted in a set of n=350 unique adjectives in 19 decades, with 19 total measures (4 measures of semantic richness, 14 measures of network centrality, and 1 measure of Zipf frequency) for each instance of a word in a decade.

We then applied GMC to analyze the lagged dependencies between these properties. Specifically, for a given measure Y_i in decade t_k, we computed the pairwise value $\mathrm{GMC}(Y_i \mid X_j)$ for all measures X_j, across all decades in the range $t_k, t_{k-1}, ..., t_{k-18}$. This approach allowed us to measure the directional dependency of Y_i on X_j over all intervals. For example, the effect of *degree* on *spectral diversity* in 2000 was computed for every measure of degree in the range 2000,1990,...1820. In addition, we computed all relations for a variable X_j in decade t_k to X_j in decades $t_{k-1},\ldots,t_{1820}$, to test the lagged effect of a variable in predicting itself.

For each GMC calculation of a measure X_j in predicting Y_i for some decade lag $0,\ldots,n$, we conducted a hypothesis test to determine whether there was a significant asymmetry in X_j predicting Y_i, as opposed to Y_i predicting X_j. We tested all possible pairwise relations for measures between groups, across all decades 1820...2000. In computing all possible between group and autocorrelated relations, we obtained n=2,500 hypothesis tests. Additionally, we applied a strict Bonferonni correction ($\alpha = 0.05/2500$). Due to the potential spuriousness of results for many individual pairwise comparisons, we do not consider results outside this significance threshold.

To further confirm the results of our GMC analysis, we used lagged correlations between measures to assess the extent to which a general temporal asymmetry is present for a given relationship (X_t, Y_{t+k}). To test this, we used a method inspired by Piazza et al. (2019) for identifying directionality in correlation data. We first calculated correlations across the full range of lag values ($k \in \{0,1,\ldots,18\}$), correlating (X_{t-k}, Y_t) and (X_t, Y_{t-k}) for each lag k. For each correlation, we grouped data by word and arranged by decade to align the time series, shifting the X and Y variables accordingly. To assess the robustness of these correlations, we applied bootstrapping with 500 iterations per lag to estimate mean correlations and 95% confidence intervals, providing a direct comparison of forward and backward lags for any X, Y relation. While Spearman's ρ only assesses the monotonic relationship between variables and does not measure the effect of conditioning a response variable on a predictor in a prior decade, the analysis allows us to test general temporal asymmetry between measures.

Results

Four measures met the criteria for temporal precedence established in our GMC analysis, either as predictor or response variables: two measures of semantic richness(spectral diversity and entropy), one measure of network centrality (betweenness) and word frequency. Across all lags, network centrality (betweenness) was a significant predictor of both semantic richness measures. This relationship was strongest at the time lag of 100 years, where the predictive effect of betweenness centrality on spectral diversity was the strongest overall effect in our results ($\beta = 0.39, p < \frac{0.05}{2500}$). Additionally, betweenness centrality predicted word frequency in all lags, although the direction of the effect changes in the longest interval, where the relationship was positive ($\beta = 0.24, p < \frac{0.05}{2500}$).

Betweenness centrality also predicted different directions in the measures of semantic richness: entropy and spectral diversity. Entropy, which measures the unevenness in the distribution across multiple senses, was associated with earlier increases in betweenness centrality. Spectral diversity, which measures the range of a word's distribution of senses, was predicted negatively by the same variable. While betweenness centrality remained a predictor of later change in semantic richness and frequency in all lags, we found no consistent relationship between semantic richness and word frequency, with only one observed effect in frequency predicting semantic richness ($\beta = -0.06, p < \frac{0.05}{2500}$). Overall, our results suggest that betweenness centrality emerges as an important factor that drives changes in word frequency and semantic richness over time. The full set of directional relationships and their effect sizes can be seen in Figure 2.

The results of the bootstrapped correlation analysis suggest a temporal asymmetry in the relationship for betweenness centrality and both entropy and spectral diversity, indicated by stronger correlations between measures in negative

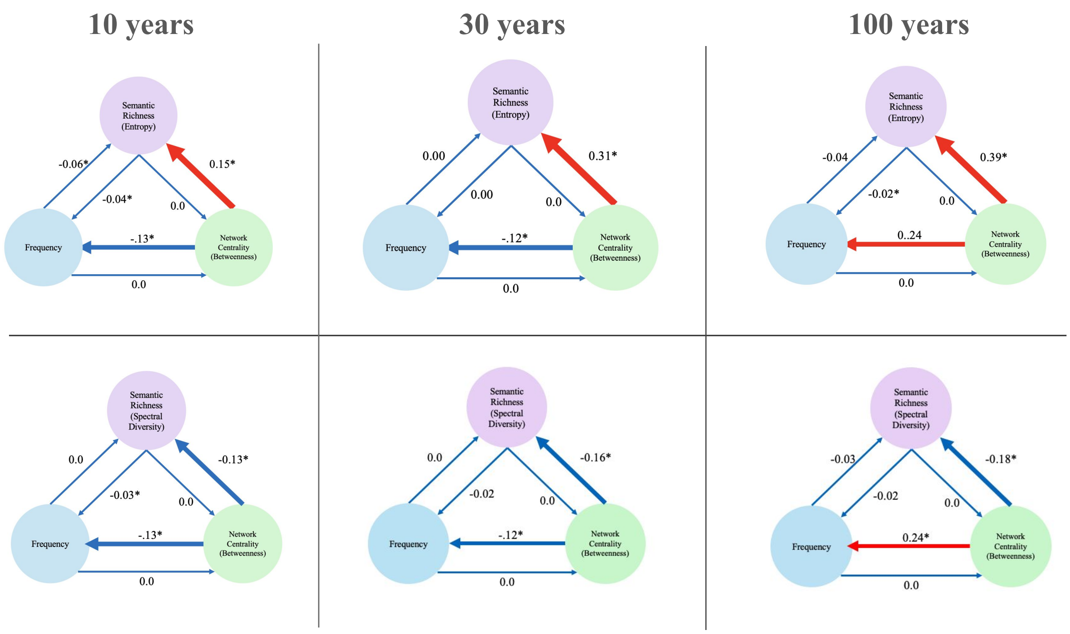

Figure 2: Changes in the dynamics of word properties over time. A direction in the graph indicates that the origin measure predicts the target measure in a later decade in GMC analysis. Effect sizes are averaged across decades. Measures marked with an asterisk (*) are robust after a Bonferonni correction of $\alpha = (0.05/2500)$. Relations to the semantic richness measure *Entropy* are pictured on the top row, and relations to the semantic richness measure spectral diversity are pictured on the bottom.

decade lags (betweenness centrality in earlier decade relative to entropy or spectral diversity) compared to later ones. In contrast, the lagged correlations for (*spectral diversity*, *frequency*) and (*frequency*, *entropy*) show no precedence for earlier decade lags, indicated by symmetry in correlations around a lag of 0, as seen in Figure 3. The relationship for betweenness centrality and word frequency shows weak temporal asymmetry in early negative decade lags (betweenness centrality preceding word frequency), and dissipates after approximately six decades. Overall, the lagged correlations indicate a precedence for earlier values of betweenness centrality in predicting measures of semantic richness (spectral diversity, entropy), an asymmetry not seen in relationships for word frequency and semantic richness. This is consistent with the results of our GMC analysis, providing additional support for an effect of betweenness centrality on semantic richness.

Discussion

How do words usage and meanings evolve within the semantic ecosystem of language? We develop a longitudinal corpus analysis to investigate the complex interplay between frequency, semantic richness, and network centrality over time. Importantly, most relationships between these variables correlated without consistent temporal asymmetry. However, among the many relationships tested, one pattern emerged consistently: a word's betweenness centrality in the semantic network predicted its future semantic richness and frequency. These relationships survived stringent statistical corrections and remained robust across different time intervals. Additionally, the relationship between network centrality and semantic richness is further corroborated by general lagged correlations between the two measures.

Notably absent was any directional predictive relationship between word frequency and semantic richness, despite previous research suggesting strong connections between these variables (Harmon & Kapatsinski, 2017; Piantadosi et al., 2012). Instead, our results suggest that betweenness centrality might mediate the relationship between frequency and meaning, though the precise mechanisms require further investigation. The consistent predictive relationship between betweenness centrality and later changes in both semantic richness and frequency suggests that a word's position within the semantic network plays a crucial role in shaping its evolutionary trajectory. Words with high betweenness centrality serve as bridges between different semantic domains. This bridging position exposes the word to increasingly diverse contexts, creating opportunities for novel uses and meanings to emerge through processes of metaphorical extension and semantic change. As these new uses become established, they may in turn lead to higher frequency as the word becomes a go-to choice for expressing similar concepts.

The increasing effect sizes over longer time intervals suggest that semantic evolution operates through gradual, cumulative processes. Small advantages conferred by a word's network position may compound over time, leading to pro-

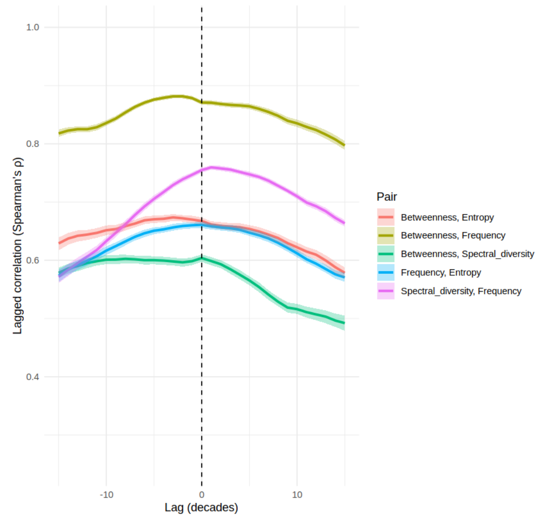

Figure 3: Bootstrapped correlations for significant GMC relationships, sampled at each lag interval. Greater correlations in the negative region of the x-axis indicate that lagged values of the first measure have greater correlation with non-lagged or later values. Conversely, correlations symmetric around a lag of 0 indicate earlier values of a measure are not predictive of later values of a measure, suggesting a lack of temporal asymmetry. All correlations are computed using Spearman's ρ.

gressively larger differences in both meaning and usage. This pattern aligns with empirical studies of semantic change that demonstrate how word meanings evolve gradually through accumulated shifts in usage patterns (Hamilton et al., 2016; Ramiro et al., 2018). As words in central network positions are used in increasingly diverse contexts, their meanings systematically expand and shift over time, leading to lasting changes in both semantics and frequency.

Although both spectral diversity and entropy capture aspects of semantic richness, we observed a consistent negative relationship between betweenness centrality and spectral diversity, in contrast to the positive relationship with entropy. This finding contradicts our initial expectation that more central words would exhibit greater semantic spread across embedding dimensions. One potential explanation for this finding is that higher centrality words develop a broad set of closely related meanings, where usage is shared across these senses (higher entropy), but these senses occupy a smaller region of semantic space (lower spectral diversity). The exact cause of this divergence remains unknown, and future work could employ additional measures of semantic spread in contextual embeddings–such as the convex-hull volume– to test whether the overall meaning space is more compact in higher betweenness centrality words.

Intriguingly, the relationship between betweenness centrality and frequency shows a complex temporal pattern: high betweenness predicts lower frequency in the shorter term (10-30 years) but higher frequency over longer periods (100 years). This reversal may reflect the processes through which novel word uses become conventionalized. Initially, words in bridging positions may be used in more experimental or marked ways, potentially reducing their overall frequency. However, as these novel uses become established over longer time periods, the word's versatility in bridging different semantic domains may lead to increased frequency much later. This pattern aligns with theories of semantic change that emphasize the gradual conventionalization of novel meanings.

However, several important limitations should be noted. While our analysis reveals temporal relationships, establishing true causality in complex linguistic systems remains challenging. Our use of decade-level time intervals may obscure more rapid semantic changes occurring at shorter timescales, particularly given the potential for fast-paced language evolution. Additionally, our semantic networks, extracted from written text, may not perfectly reflect the cognitive organization of meanings in speakers' minds—though this limitation is inherent to historical linguistic research where we cannot directly access past speakers' mental representations. Our focus specifically on adjectives means these patterns may not generalize to other word classes—verbs and nouns, for instance, might show different evolutionary dynamics given their distinct semantic and syntactic properties.

Overall, our results suggest that a word's position on the lexical workbench—its centrality within the semantic network—might play a crucial role in determining its evolutionary trajectory. Like well-placed tools that become increasingly versatile through creative repurposing, words in bridging positions appear particularly prone to semantic extension and increased usage. Understanding these network effects could help explain why certain words develop diverse meanings while others remain specialized, why certain words pick up in frequency while some fall out of the use, and might even allow us to predict which words are most likely to undergo semantic change and change in frequency in the future.

References

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, *16*(1), 22–29.

Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in english second language speakers. *Language Learning*, *60*(3), 573–605.

Cruse, D. A. (1986). Lexical semantics. *Cambridge University*.

Darwin, C. (1872). *The descent of man, and selection in relation to sex*. John Murray.

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and plants: A study in coevolution. *Evolution*, 586–608.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in cognitive sciences*, *23*(5), 389–407.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, *2016*, 2116–2121.

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

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing* [To appear].

Liu, Q., De Deyne, S., & Lupyan, G. (2024). Why are some words more frequent than others? new insights from network science.

Liu, Q., De Deyne, S., Jiang, X., & Lupyan, G. (2023). Understanding the frequency of a word by its associates: A network perspective. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45).

Mahowald, K., Dautriche, I., Gibson, E., & Piantadosi, S. T. (2018). Word forms are structured for efficient use. *Cognitive science*, *42*(8), 3116–3134.

Müller, F. M. (1870). Darwinism tested by the science of language. translated from the german of professor august schleicher, by dr. alex, v. w. bikkers [Original source of the "struggle for life" quote, as cited by Darwin in *The Descent of Man*]. *Nature*, 257.

Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291.

Piazza, E. A., Hasenfratz, L., Hasson, U., & Lew-Williams, C. (2019). Infant and adult brains are coupled to the dynamics of natural communication. *Psychological Science*, *30*(6), 822–830.

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*, *115*(10), 2323–2328.

Schleicher, A. (circa 1860). On the evolution of language [A pamphlet originally written in German. This work, which introduced an early analogy of language evolution, was translated by Dr. Alex, V. W. Bikkers and reviewed by Max Müller in *Nature* (1870).].

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.

Thompson, J. N. (1994). *The coevolutionary process*. University of Chicago press.

Trott, S., & Bergen, B. (2022). Languages are efficient, but for whom? *Cognition*, *225*, 105094.

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.

Zheng, S., Shi, N.-Z., & Zhang, Z. (2012). Generalized measures of correlation for asymmetry, nonlinearity, and beyond. *Journal of the American Statistical Association*, *107*(499), 1239–1252.

Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of general psychology*, *33*(2), 251–256.

Zipf, G. K. (1949). Human behavior and the principle of least effort: An introduction to human eoclogy.