Cognitive Factors in Word Sense Decline

Aniket Kali
Department of Computer Science
University of Toronto
(aniket.kali@mail.utoronto.ca)

Yang Xu
Department of Computer Science
Cognitive Science Program
University of Toronto
(yangxu@cs.toronto.edu)

Suzanne Stevenson
Department of Computer Science
University of Toronto
(suzanne@cs.toronto.edu)

Abstract

Word senses rise and fall due to a variety of causes. Previous research has explored how words grow novel senses, but the opposite problem of word sense decline is much less studied. Inspired by recent work on word decline, we investigate the cognitive factors that might explain the historical decline of word *senses*. We formalize a set of eight psycholinguistic predictors and assess their roles in discriminating declining senses from stable ones over the past two centuries in English. We find that semantic density, change in usage frequency in the semantic neighbourhood, and contextual diversity all predict word sense decline. Our study elucidates the cognitive underpinnings of word sense decline as the lexicon evolves.

Keywords: lexicon; word meaning; psycholinguistic properties; lexical semantic change; word sense decline

Introduction

Words often take on new senses, but the opposite phenomenon of word sense decline is also commonly observed. For example, the word *language* has gradually seen a decline in its sense of "style of writing or speech" (e.g., "He was used to the obscure and euphemistic language of legal documents"). In contrast, the word *marriage* continues to be used in its sense of a "formal union of two partners in a personal relationship" (e.g., "She thought her marriage was the tragedy of her life"); see Figure 1. The decline of word senses may be influenced by external factors driven by social, technological, or cultural changes (e.g., the sense of "religious garment" for the word *habit* has declined, plausibly due to increased secularization). Here we focus on understanding whether there are language-internal factors, driven by cognitive constraints, that might explain why certain word senses decline while others remain stable, thereby exploring the cognitive underpinnings of word sense decline in the evolution of the lexicon.

Prior work has suggested that for words, the processes of birth and decline are often not arbitrary but rather are constrained by cognitive and communicative principles (Ryskina et al., 2020; Francis et al., 2021). Using quantitative tools, these studies examine how semantic, distributional, and phonological factors predict the birth and decline (respectively) of words in the English lexicon. For example, factors related to the supply and demand of words in the semantic space (i.e., whether they fill an emerging communicative need) may predict where new words are likely to arise (Ryskina et al., 2020). In studying the decline of words, Francis et al. (2021) conceptualize the lexicon as an ecosystem: words which are less

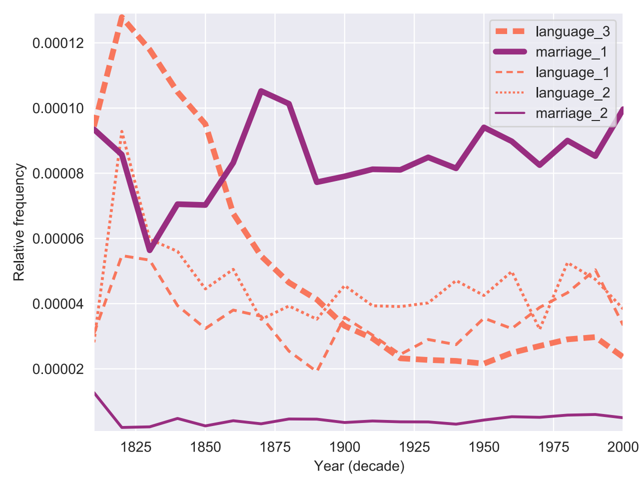

Figure 1: An illustration of a declining and stable pair of senses in our study: `language_3` ("style of writing or speech") and `marriage_1` ("formal union of two partners in a personal relationship"). Thinner lines with the same colour are other senses of the same word. These include: `language_1` ("the method of human communication"), `language_2` ("a communication system used by a particular community"), and `marriage_2` ("a mixture of elements").

cognitively advantageous are less ecologically fit, and hence they are prone to decline out over time. Their idea is grounded in psycholinguistic work suggesting that certain properties may make a word more or less difficult to learn or mentally retrieve; these are formalized into a set of cognitive factors used to discriminate historically declining words from stable words. Francis et al. (2021) find a role for semantic and distributional factors, such as neighbourhood semantic density (Chen & Mirman, 2012) and contextual diversity (Adelman et al., 2006; Stewart & Eisenstein, 2018), suggesting how these specific cognitive factors predict lexical decline. Related research has also identified similar factors relevant to synonym competition and more broadly, lexical competition (Baumann et al., 2023; Karjus et al., 2020b; Li et al., 2024; Turney & Mohammad, 2019). Our work extends these studies to examine decline at the fine-grained level of word senses.

These patterns in lexical dynamics fit into a broader theme of regularity in lexical semantic change (e.g., Traugott &

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Dasher, 2001), where others have identified factors such as word frequency and degree of polysemy (i.e., how many senses a word has) to account for the rates of semantic change and loss (e.g., Hamilton et al., 2016; Luo & Xu, 2018). Building on this line of research, we hypothesize that there are cognitive factors that predict the fate of word senses, such that they can reliably discriminate declining senses from stable senses through time. To our knowledge, this is an under-explored area in the cognitive science of language and language change.

Our study complements an extensive line of work exploring cognitive principles in the growth and acquisition of word meaning. For example, prior work has shown that novel senses tend to emerge in words with closely related senses (Ramiro et al., 2018), and word senses grow in systematic ways (Fugikawa et al., 2023; Xu et al., 2017). These principles are reflected in language development, where children extend known words to novel referents in predictable ways (e.g., Clark, 1978; Rescorla, 1980) and tend to use polysemy to facilitate their acquisition of new word meanings (Floyd & Goldberg, 2021; Srinivasan & Rabagliati, 2021). Our focus is to investigate word sense decline, which can be considered as the opposite problem of word meaning growth. We believe that the processes of decline may be fundamentally constrained and understood on similar bases of human cognition.

Inspired by the framework of Francis et al. (2021) on word decline, we approach the problem of sense decline by formalizing a set of cognitive factors using quantitative methods from distributional semantics and data drawn from large-scale psychological experiments. Using these factors, we compare declining and stable senses through a 200-year historical period of English and assess which factors play a key role in predicting the declining senses from the stable senses; Figure 1 gives examples from our data. To preview our results, we find that neighbourhood semantic density, change in neighbourhood frequency, and diversity of usage contexts are key predictors, suggesting these are some of the key cognitive influences on word sense decline.

Cognitive Factors

We describe a set of cognitive factors belonging to three thematic categories, based on the different roles they play in cognitive processing of word senses. First, we know that the semantic neighbourhood of a sense – the related senses around it in semantic space – can give rise to both competitive and cooperative effects on linguistic processing (Armstrong & Plaut, 2016; Chen & Mirman, 2012; Srinivasan & Rabagliati, 2021). The underlying **meaning structure** of a word sense's neighbourhood may aid its processing in lexical access. Second, **semantic features**, such as the level of concreteness or polarity (valence), can affect lexical retrieval in the long term (Luo et al., 2019; Vejdemo & Hörberg, 2016), and might contribute to a sense's functional value. Third, senses can be used narrowly or broadly, and frequently or infrequently, thus varying in **usage and distribution**. This can lead to increased or decreased entrenchment in the lexicon, which has been shown to influence word survival (Balota & Spieler, 1999; Stewart & Eisenstein, 2018).

We select factors that were motivated from a lexical processing perspective, and that could be assessed automatically (for large-scale study). Specifically, we devise factors that could predict, based on initial conditions at a historical time ($t = 0$), a substantial decline in sense usage, by a terminal timepoint ($t = n$). Our list of factors and their predicted direction of correlation with decline are outlined in Table 1.

Meaning Structure

First, we consider the semantic neighbourhood; if it is more crowded, multiple senses may compete to express the same meaning. Increased **semantic density** at the word level is known to have an inhibitory effect on production and recognition due to increased competition among nearby meanings (Chen & Mirman, 2012; Mirman & Magnuson, 2008). We hypothesize this may influence decline at the sense level, just as Francis et al. (2021) observed that greater semantic density correlates with decline at the word level.[1]

Ryskina et al. (2020) found that new words emerge in increasingly popular neighbourhoods (i.e., increasing frequency over time), reflecting the relevance of the changing importance of topics in discourse. Related to this, we expect senses in neighbourhoods of falling **demand** may be more likely to decline, belonging to discourse topics of decreasing popularity.

Experimental studies have found that polysemous senses of a word activate and reinforce each other in memory, due to their overlapping meanings (e.g., Floyd and Goldberg, 2021; Srinivasan and Rabagliati, 2021). However, this "polysemy advantage" is a graded effect: more peripheral (less related) senses may not benefit from mutual activation (and may even experience competitive effects) (Armstrong & Plaut, 2016; Brocher et al., 2016; Klepousniotou et al., 2008; Rodd et al., 2002). We hypothesize then that increased **sense peripherality** to other senses of the same word may be associated with decline. Likewise, senses belonging to words with a fewer **number of senses** may (all else being equal) be less likely to be indirectly activated by the other senses of the word, and thus more likely to decline.

Semantic Features

Meanings considered concrete (i.e., physically tangible) are more easily retained than abstract meanings (e.g., De Groot and Keijzer, 2000; Vejdemo and Hörberg, 2016). Lower **concreteness** may then make a sense more difficult to retrieve, leading to its decline.

Words also shift in valence or sentiment (e.g., Cook and Stevenson, 2010; Ullmann, 1962). For instance, metaphorical meaning extension can shift the meaning of a word from neutral to more polarized sentiment (Xu et al., 2017). On the other hand, communicative processes like the creation of euphemisms indicate pressures to use less polarized words

[1] Complementarily, Ryskina et al. (2020) found that lower semantic density facilitated the rise of new words, due to a lack of 'supply' of needed words in a semantic region.

Category	Factor	Pred.
meaning structure	semantic density	+
	demand	−
	sense peripherality	+
	number of senses	−
semantic features	concreteness	−
	valence	+
usage and distribution	word frequency	−
	contextual diversity	−

Table 1: The list of 8 factors organized by category. A positive (+) prediction indicates an expectation that as the factor increases, the sense declines, whereas negative (−) indicates the opposite.

(Burridge, 2012). We hypothesize that senses with strong **valence** (i.e., a strongly positive or negative connotation) may fall out of use over time.

Usage and Distribution

Higher **word frequency** has been shown to lead to greater entrenchment in the lexicon (e.g., Balota and Spieler, 1999; Murray and Forster, 2004). We hypothesize that a word with lower initial frequency (i.e., in the 1810s) provides fewer opportunities for any of its senses to be (directly or indirectly) activated, leading to less reinforcement and possible decline.

Others have argued that a focus on word frequency is a limited view of lexical entrenchment. Instead, usage in a wide variety of contexts permits a word to have many "niches" within the lexicon, ensuring entrenchment for long-term survival (e.g., Adelman et al., 2006; Stewart and Eisenstein, 2018). A lack of **contextual diversity** may then leave a sense prone to fall out of use (as shown for words by Francis et al., 2021).

Materials and Methods

Central to our analysis is comparing the identified cognitive factors across a set of declining and stable senses (cf. Francis et al., 2021; Ryskina et al., 2020). To identify declining and stable senses, we need a historical corpus for which we can determine the senses of a (large) set of word usages, such that we can calculate counts of word senses (i.e., not just counts of words) and determine which senses decline in frequency over time, and which are relatively stable across the historical period. We begin by describing the data used in this study, and how we obtained temporal sense counts. Next, we outline how we select the declining and stable senses to study, as well as how we match them for comparison. Finally, we provide detailed operationalizations for each of our 8 factors.[2]

Data and Sense Classification

We use the Clean Corpus of Historical American English (CCOHA; Alatrash et al. 2020; Davies 2012), constructed from a balanced variety of sources, including fiction, newspapers, magazines, and so on. It spans the period 1810–2009, which we study in full, removing duplicate occurrences of a usage, and aggregating data per-decade (i.e., the 1810s is 1810–1819).

Since this corpus is not sense-tagged, we use the automatic method of Hu et al. (2019) to do so. This method creates a "ground-truth" sense representation for each sense of a word given in a dictionary, using the example sentences from the dictionary listing for the word. That is, the examples in the dictionary serve as labelled data from which to create a representation of each sense of a word. Each untagged usage of the same word in a corpus can be compared to these sense representations, tagging that usage with the most similar sense.

Hu et al. (2019) operationalize their approach by using the pre-trained `bert-base-uncased` version of BERT (Devlin et al., 2019), a popular language model used to create semantic representations. Given a usage of a target word, BERT returns a contextualized usage embedding: a unique 768-dimensional vector computed from the word and its context. Hu et al. (2019) compute such embeddings for 10 of the example usages accompanying a sense, drawn from a version of the Oxford English Dictionary. They then compute the mean of these dictionary-example embeddings to create the ground-truth sense embedding – a 768-dimensional vector which represents the given sense. We obtained their set of sense embeddings for this study, totalling 15836 senses from 3220 words.[3]

For this set of words, we computed contextualized usage embeddings for all usages in CCOHA, also using the pre-trained `bert-base-uncased` version of BERT. We compared each usage embedding of a word to the sense embeddings of that same word, tagging each corpus usage with the label of the sense that produced the maximum cosine similarity (matching for part-of-speech, as in Hu et al., 2019).[4] To obtain sense counts per decade, as shown in Figure 1, we then counted the number of usages tagged with a given sense in each decade, and normalized this value using the total number of usages (across all words) in that same decade.[5] Finally, we removed any sense which had an initial normalized frequency below 1/1M, leaving us with a final set of 8430 senses.

Identifying Declining and Stable Senses

A declining sense is one which observes a gradual decline to near-zero usage, followed by a period of infrequent usage

[2]The code repository is available here: https://github.com/an-k45/cognitive-sense-decline.

[3]This set of 3220 words covers approximately 75% of all individual usages in CCOHA, in any given decade. 20 words had more usages than we could reasonably compute embeddings for, so we downsampled to 2M usages and then used their raw total count.

[4]Despite being trained on Modern English, BERT has been used successfully to analyze historical data (e.g., Giulianelli et al., 2020), although there may be limitations (e.g., Hoeken et al., 2023).

[5]Hu et al. (2019) normalize their per-decade counts using only the usage count of the word in question, rather than usages across all words. That approach reveals word-internal dynamics, but does not accurately capture the phenomenon of sense decline, since a sense that is declining in its usage may appear stable or even increasing relative to the rise and fall of other senses of the same word.

word	sense definition	`sem_dens`	`demand`	`c_div`
D: position	"A proposition laid down or asserted; a tenet or assertion."	**0.70**	$-$**0.18**	**0.27**
S: business	"A person's regular occupation, profession, or trade."	**0.66**	**0.05**	**0.36**
D: occupy	"Fill or preoccupy (the mind)"	0.67	$-$**0.44**	0.51
S: repair	"Restore (something damaged, faulty, or worn) to a good condition."	0.72	$-$**0.18**	0.40

Table 2: Examples of matched declining ('D') and stable ('S') word senses with values for select factors. Bolded values indicate the difference in the factors of the matched pair (D − S) is consistent with the prediction in Table 1.

(possibly at 0). Likewise, a stable sense is one whose usage remains fairly constant.

Following Francis et al. (2021), we identify declining senses by fitting frequencies to the piecewise function defined in Eqn. (1). We apply piecewise linear regression, obtaining the curve which minimizes the Mean Square Error (MSE).

$$x(t) = \begin{cases} a(b-t) & \text{if } t \leq b \\ 0 & \text{if } t > b \end{cases} \quad (1)$$

In this equation, t is the decade, and a and b are both positive model parameters, with b defined in the range $(1,20)$. This function therefore has a declining-piece (meeting the x-axis at b), and a zero-piece, defining an 'elbow'-curve. We filtered out senses which had too high an MSE score, or did not exhibit sufficient decline between the maximum and final frequency.

Similarly for stable senses, we fit them to the horizontal line at their median frequency. Once again, we removed those senses with too high an MSE score, or whose maximum and minimum frequencies varied greatly.[6]

This process yields 855 declining and 769 stable senses, from which we next select our matched pairs of declining and stable senses.

Matching Declining and Stable Senses

To assess which of our cognitive factors predict decline, we compare declining and stable senses. We cannot make a comparison between any arbitrary pair of senses, since they may have properties which are associated with decline and co-vary with the factors we investigate. Following Francis et al. (2021), we thus adhere to a controlled evaluation paradigm by finding pairs of declining and stable senses matched on a set of potential confounds.

Specifically, we pair together a declining and stable sense if they: share a comparable initial frequency ($\pm$10%), belong to different words with a similar number of senses ($\pm$2 senses), share the same part-of-speech (noun to noun, verb to verb, etc), are not within the same semantic neighbourhood (as defined by the 10 nearest neighbouring senses[7]), and are of similar lengths ($\pm$2 characters). Additionally, the sum of the character length of all words in the set of stable senses and declining senses could differ by no more than 1, to ensure that the words for one set were not consistently longer than those for the other. This process yields a set of 412 matches (with 824 total senses).[8]

Predictive Factors of Decline

Here we describe the detailed operationalization of each of our 8 factors listed in Table 1. For most factors, we compute values for all 412 matches. In certain cases, as noted below, we compute values for only a subset of matches to ensure consistency in the cognitive phenomena that we are measuring. Missing entries were assigned a default value equal to the mean of all computed values of the same kind of sense (i.e., other declining or stable senses only) for the given factor.[9] We report example values for select factors and senses in Table 2.

Meaning Structure

Semantic Density (`sem_dens`): We measure how dense a neighbourhood is by computing the mean cosine similarity from a target sense to its 10 nearest neighbours.

Demand (`demand`): The growing popularity, or demand, of a semantic neighbourhood is reflected by the increasing frequency of its members. We formalize this idea as Ryskina et al. (2020) did, computing the Spearman correlation between the ordered sequence of decades $\{1,2,...,20\}$ and the ordered frequencies for those decades $f_{(1:20)}(n_i)$, for each sense n_i in the target sense's neighbourhood. In this way, we compute the extent to which we see consistent and growing popularity for the semantics of a sense. Then, to obtain the change in neighbourhood demand, we report the mean correlation for the 10 nearest neighbours, as defined in Eqn. (2).

$$d = \frac{1}{10}\sum_{i=1}^{10} r_s(\{1,2,...,20\}, f_{(1:20)}(n_i)) \quad (2)$$

This measure of demand relies on consistent change, but decline may start at any point (i.e., declining senses may initially remain stable or even rise). Given this, we assigned default values for any pair of senses whose declining sense

[6]For both the declining and stable sets, if more than one of the identified senses belongs to the same word, we retained just one of them (to avoid word-based confounds); we kept the sense with the highest initial frequency, to maximize amount of data.

[7]Our results were robust to using 20 or 50 neighbours.

[8]During the matching process, many senses had multiple eligible matches. We present the results using one set of pairs, but found our results were robust to various possible sets of matches.

[9]We also took the approaches of (1) dropping all matches with any missing values, leaving a smaller data set, or (2) removing all filters, ensuring all matches had values for every factor. In both cases, results were largely consistent with using mean imputation.

began to fall well after the initial timepoint, as indicated by fit to the 'elbow'-curve. This results in 83 out of 412 matches being assigned a default value.[10]

Sense Peripherality (`s_perph`): Given a sense s_i, we measure peripherality as the mean cosine distance from s_i to all other senses s_j belonging to the same word. Senses from words with no other senses are assigned the default value, since they have no fellow senses to be peripheral to. In total, 24 declining senses, and 37 stable senses, out of 412 total senses in each set, are assigned a default value.

Number of Senses (`s_num`): For a given sense, this is the number of distinct senses in the dictionary used by Hu et al. (2019) and met our initial frequency threshold.

Semantic Features

We use available **concreteness** and **valence** ratings of *words* to infer ratings of our target *senses*. We address two issues: not all words of our senses have ratings, and word-level ratings may not be accurate (i.e., not all senses of a word necessarily have the same level of concreteness or valence; cf. Rodd et al., 2002). We use a two-step process. We first find a word-based rating for all senses in our corpus. For senses whose words lack a rating in the resource, we impute these values using a linear regression model on known ratings with sense embeddings as predictors. In the next step, we use these model-predicted word-based ratings to find a sense-based rating: We compute the rating for each of our target senses as an average over the word-based ratings of its (sense) neighbours, assuming this mean rating may be more nuanced than the target sense's simple word-based rating.

Concreteness (`conc`): We use the resource of Brysbaert et al. (2014), a set of 40,000 word-level human ratings of concreteness. We center ratings at 0 to range from -2 (highly abstract) to 2 (highly concrete). We compute concreteness of a sense as the mean (predicted) concreteness value of its 10 nearest neighbours.

Valence (`val`): We use the resource of Warriner et al. (2013), a set of 13,915 word-level human ratings of valence. We center ratings at 0 to range from -4 (highly negative) to 4 (highly positive) then take the absolute value. We compute valence of a sense as the mean (predicted) valence value of its 10 nearest neighbours.

Usage and Distribution

Word Frequency (`w_freq`): Since our predictions about sense decline are based on conditions at the initial timepoint, we report the frequency of the word in the 1810s.

[10]We also applied tighter cut-offs for fit, which set default values for a greater number of pairs, and found the results were robust.

factor	dec	stb	diff.
sem_dens**	0.68 (±0.05)	0.70 (±0.04)	− ✗
demand**	−0.35 (±0.26)	−0.06 (±0.29)	− ✓
s_perph*	0.22 (±0.09)	0.19 (±0.08)	+ ✓
s_num	3.33 (±1.72)	3.21 (±1.72)	+
conc**	−0.09 (±0.38)	0.06 (±0.43)	− ✓
val	0.53 (±0.46)	0.50 (±0.40)	+
w_freq	1.19 (±1.09)	1.09 (±0.96)	+
c_div*	0.37 (±0.06)	0.39 (±0.06)	− ✓

Table 3: Mean (±SD) for each factor, for both sets of senses; $\texttt{w\_freq} \cdot 10^4$ is shown. 'diff.' indicates the direction of the difference correlating with decline, with a ✓, or ✗, for predictions which do, or do not, match those in Table 1 respectively. Significant factors are bolded, and marked as '*' ($p < 0.01$) or '**' ($p < 0.001$), after a Bonferroni correction.

Contextual Diversity (`c_div`): The contextual diversity of a sense is understood as the range of contexts it is used in. Given that usage embeddings are computed based on context, the greater the range of these embeddings in semantic space, the higher the sense's contextual diversity. Specifically, we compute the average pairwise cosine distance between all usage embeddings of a sense in the 1810s.

Results

We compute the values of the 8 cognitive factors over the set of declining senses (`dec`) and stable senses (`stb`). We first examine whether our sets of paired senses (`dec` and `stb`) show a significant difference for each factor, and then assess which factors stand out in their ability to predict decline when considered in combination.

Analysis of Individual Factors

We begin by examining the predictive ability of each factor on its own. Specifically, we apply the Wilcoxon signed-rank test (a non-parametric paired test) to assess whether the sets of (paired) senses `dec` and `stb` are significantly different.

We report our findings in Table 3. In total, 5 of our 8 factors were associated with significant differences between the sets `dec` and `stb`: `sem_dens`, `demand`, `s_perph`, `conc`, and `c_div`. Further, by taking the difference between the reported mean values (`dec` − `stb`), we find that the predicted directions in Table 1 hold for all significant factors, except in the case of `sem_dens`. For the latter, contrary to our prediction, a denser neighbourhood is associated with survival of a sense; we return to this in our discussion.

Given that our factors may overlap in what cognitive phenomena they measure, we compute the correlation between the differences of the 412 sense pairs, for each factor; see Figure 2. Among the significant factors from above, `sem_dens` is correlated with `s_perph`, and `demand` is correlated with `conc`, showing that each pair may tap into similar conditions associated with decline.

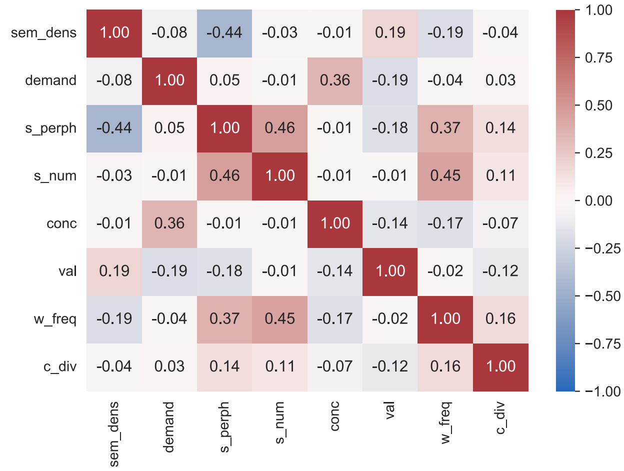

Figure 2: Heatmap of Pearson correlations for each factor, using differences between every pair of senses.

pred.	β coef.	std err	z	p
const	10.36	1.59	6.54	0.00
sem_dens	−6.39	1.28	−5.00	0.00
demand	−11.19	1.24	−9.02	0.00
s_perph	1.00	1.24	0.81	0.42
s_num	1.29	1.38	0.93	0.35
conc	−1.58	1.00	−1.58	0.11
val	−0.41	0.90	−0.46	0.65
w_freq	−0.73	1.48	−0.50	0.62
c_div	−3.08	1.01	−3.06	0.00

Table 4: Summary of the logistic regression model used to predict direction (1 for dec − stb, 0 for stb − dec). Significant predictors are bolded.

Prediction of Decline from Multiple Factors

To consider the cognitive factors in a joint setting, we use a logistic regression model to predict decline by combining these factors. Given that our individual senses are paired and therefore not independent, our predictions are over the set of 412 sense pairs given each of our 8 factors. In the logistic regression, we can predict 1 for a random half of pairs ordered as dec:stb, and 0 for the for the other half in reverse order stb:dec, as the dependent variable. We compute 8 independent variables, corresponding to each of our factors, by taking dec − stb or stb − dec based on the assigned order (and scale all values to the 0–1 range).

The logistic regression model yields a pseudo-r^2 value of 0.43, and a leave-one-out cross validation test achieves an 82% accuracy rate. Table 4 outlines the results of our model in further detail.[11] Now, just three factors come out as significant: sem_dens, demand, and c_div. This outcome is sensible: the two factors which are no longer significant, s_perph and conc, were fairly strongly correlated with other factors which are significant here. We observe that demand and c_div have negative β coefficients, consistent with our prediction that they should decrease with decline. Once again, sem_dens comes out as opposite to our prediction: its negative β coefficient means a sense is more likely to decline if semantic density is lower.

Discussion and Conclusion

We offer a quantitative analysis of word sense decline from a cognitive perspective, approaching this problem by proposing a set of relevant cognitive factors and testing their ability to predict declining and stable senses.

Our findings highlight three important factors. Lower contextual diversity predicts sense decline, just as Francis et al. (2021) predicted for word decline. A narrower range of word usages indicates weak lexical entrenchment, now found for both the word and sense levels (Adelman et al., 2006; Stewart & Eisenstein, 2018). Likewise, lower demand increases the likelihood of decline. As a semantic neighbourhood becomes more obscure, its members may become harder to recall and lack the same importance in discourse. This complements the findings of Ryskina et al. (2020) that upstart neighbourhoods are "stylish" and facilitate the rise of new words.

Finally, we find that sense decline is associated with lower semantic density (greater sparsity). At the word level, semantic density and sparsity facilitate the decline and rise of words, respectively (Francis et al., 2021; Ryskina et al., 2020). This is consistent with the effect of competition between words in a dense neighbourhood, and cooperation, which facilitates processing in a sparse neighbourhood (Chen & Mirman, 2012; Mirman & Magnuson, 2008). However, Karjus et al. (2020a) notes how competition (and decline) between near-synonyms occurs only for words belonging to a semantic topic of static importance over time; increasingly important topics permit co-existence between near-synonyms, indicating the outcome of cognitive factors may vary at different levels of meaning.

One explanation might be that competition causing decline amongst similar words is preferable as it simplifies the lexicon. The decreased relevance of this at the sense level could result in a weaker competition effect, increasing the relative importance of the cooperation effect (Buchanan et al., 2001; Chen & Mirman, 2012). Perhaps like the "polysemy advantage" (Floyd & Goldberg, 2021; Srinivasan & Rabagliati, 2021), greater local density among senses aids processing and retention by virtue of mutual activation and reinforcement–though the precise manner of this remains a topic for future study. In addition, our predictions drew from initial conditions, and diachronic analysis might further understanding of how factors interact to influence decline.

We have shown that the dynamics of word senses are subject to cognitive constraints placed on the lexicon, which reliably predict whether a given sense will decline or remain stable. In doing so, we contribute to a broader understanding of lexical dynamics at the sense level grounded in human cognition.

[11] We find consistent results across 10 random shuffles of the data.

References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychol. Sci.*, *17*(9), 814–823.

Alatrash, R., Schlechtweg, D., Kuhn, J., & Schulte im Walde, S. (2020). CCOHA: Clean corpus of historical American English. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 6958–6966.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, *31*(7), 940–966.

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*(1), 32.

Baumann, A., Stephan, A., & Roth, B. (2023, December). Seeing through the mess: Evolutionary dynamics of lexical polysemy. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 8745–8762). Association for Computational Linguistics.

Brocher, A., Foraker, S., & Koenig, J.-P. (2016). Processing of irregular polysemes in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1798.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behav. Res. Methods*, *46*(3), 904–911.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*(3), 531–544.

Burridge, K. (2012). Euphemism and language change: The sixth and seventh ages. *Lexis. Journal in English Lexicology*, (7).

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychol. Rev.*, *119*(2), 417–430.

Clark, E. V. (1978). Strategies for communicating. *Child Development*, 953–959.

Cook, P., & Stevenson, S. (2010). Automatically identifying changes in the semantic orientation of words. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Davies, M. (2012). Expanding horizons in historical linguistics with the 400-million word corpus of historical american english. *Corpora*, *7*(2), 121–157.

De Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*(1), 1–56.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (pp. 4171–4186). Association for Computational Linguistics.

Floyd, S., & Goldberg, A. E. (2021). Children make use of relationships across meanings in word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 29–44.

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

Fugikawa, O., Hayman, O., Liu, R., Yu, L., Brochhagen, T., & Xu, Y. (2023). A computational analysis of crosslinguistic regularity in semantic change. *Frontiers in Communication*, *8*, 1136338.

Giulianelli, M., Del Tredici, M., & Fernández, R. (2020). Analysing lexical semantic change with contextualised word representations. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 3960–3973). Association for Computational Linguistics.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1489–1501). Association for Computational Linguistics.

Hoeken, S., Alacam, Ö., Fokkens, A., & Sommerauer, P. (2023, December). Methodological insights in detecting subtle semantic shifts with contextualized and static language models. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the association for computational linguistics: Emnlp 2023* (pp. 3662–3675). Association for Computational Linguistics.

Hu, R., Li, S., & Liang, S. (2019). Diachronic sense modeling with deep contextualized word embeddings: An ecological view. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 3899–3908). Association for Computational Linguistics.

Karjus, A., Blythe, R. A., Kirby, S., & Smith, K. (2020a). Communicative need modulates competition in language change. *CoRR*, *abs/2006.09277*.

Karjus, A., Blythe, R. A., Kirby, S., & Smith, K. (2020b). Quantifying the dynamics of topical fluctuations in language. *Language Dynamics and Change*, *10*(1), 86–125.

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.

Li, Y., Breithaupt, F., Hills, T., Lin, Z., Chen, Y., Siew, C. S., & Hertwig, R. (2024). How cognitive selection affects language change. *Proceedings of the National Academy of Sciences*, *121*(1).

Luo, Y., Jurafsky, D., & Levin, B. (2019). From insanely jealous to insanely delicious: Computational models for the semantic bleaching of English intensifiers. In N. Tahmasebi, L. Borin, A. Jatowt, & Y. Xu (Eds.), *Proceedings of the 1st international workshop on computational approaches to historical language change* (pp. 1–13). Association for Computational Linguistics.

Luo, Y., & Xu, Y. (2018). Stability in the temporal dynamics of word meanings. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 65–79.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*(3), 721.

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.

Rescorla, L. A. (1980). Overextension in early language development. *Journal of Child Language*, *7*(2), 321–335.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of memory and language*, *46*(2), 245–266.

Ryskina, M., Rabinovich, E., Berg-Kirkpatrick, T., Mortensen, D., & Tsvetkov, Y. (2020). Where new words are born: Distributional semantic analysis of neologisms and their semantic neighborhoods. In A. Ettinger, G. Jarosz, & J. Pater (Eds.), *Proceedings of the society for computation in linguistics 2020* (pp. 367–376). Association for Computational Linguistics.

Srinivasan, M., & Rabagliati, H. (2021). The implications of polysemy for theories of word learning. *Child Development Perspectives*, *15*(3), 148–153.

Stewart, I., & Eisenstein, J. (2018). Making "fetch" happen: The influence of social and linguistic context on nonstandard word growth and decline. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 4360–4370.

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

Turney, P. D., & Mohammad, S. M. (2019). The natural selection of words: Finding the features of fitness. *PLOS ONE*, *14*(1), 1–20.

Ullmann, S. (1962). *Semantics: An introduction to the science of meaning* (1st ed.). Blackwell.

Vejdemo, S., & Hörberg, T. (2016). Semantic factors predict the rate of lexical replacement of content words. *PLOS ONE*, *11*(1), 1–15.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behav. Res. Methods*, *45*(4), 1191–1207.

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.