Why do some words have more meanings than others? A true neutral model for the meaning-frequency correlation.

Alexey Koshevoy (alexey.koshevoy@univ-amu.fr)
Laboratoire de Psychologie Cognitive, Universite Aix-Marseille, Marseille, France
Institut Jean Nicod, Département d'études cognitives, ENS-PSL, Paris, France

Isabelle Dautriche
Laboratoire de Psychologie Cognitive, Universite Aix-Marseille, Marseille, France

Olivier Morin
Institut Jean Nicod, Département d'études cognitives, ENS-PSL, Paris, France

Abstract

The lexica of natural languages are ambiguous, but the degree of ambiguity is unequal between words. Some words have more meanings than others. However, the exact properties that favor some words over others when acquiring a new meaning are not very well understood. In recent years, several studies suggested that some words gain more meanings than others based on selection for efficient communication, which could explain the correlation between meaning and frequency discovered by Zipf (Piantadosi, Tily, & Gibson, 2012; Gibson et al., 2019). The object of this study is to assess the role of selection in the meaning-frequency correlation using a neutral model that yields a meaning-frequency correlation without selection pressures. We provide a model where words gain additional meanings through reuse. In the neutral model presented in this paper, words are chosen to be reused at random, independently of their frequency, hence there is no selection mechanism favoring efficient communication. Unlike previous attempts to introduce null models of the meaning-frequency correlation (Caplan, Kodner, & Yang, 2020; Trott & Bergen, 2020), it truly does not rely on selection for frequency. We show that statistical regularities related to ambiguity, such as Zipf's meaning-frequency correlation, can arise in conditions when words are not undergoing any selective pressures. This model has the additional property of matching word frequency distributions of natural languages. It can provide the baseline against which the presence of selection for efficient communication in natural languages can be assessed.

Keywords: lexicon; Zipf meaning-frequency correlation; efficient communication; modeling; ambiguity; neutral model

Introduction

Ambiguity is a pervasive phenomenon that can be encountered on many levels in natural languages. For instance, an English word such as "bat" has several meanings. It can mean *a small flying mammal*, *a club used to play baseball* or *hitting something with a club in different sports*. On the other hand, words like *hair dryer* correspond to a unique meaning – a device for drying one's hair. Ambiguity is pervasive beyond content words: grammatical markers, such as English '-en', can be used to express causation (*enligth-en*) – *to make something clear to someone* or to form adjectives; "wood-en" – *something that is made of wood*. On the other hand, the "-er" suffix in English can only be used as an agentive nominalizer when combined with a noun; *work-er*, *lawy-er*, *driv-er*, etc. Additionally, patterns of lexical ambiguity vary cross-linguistically (François, 2008); for instance, in Russian, the same word, "d'erevo" is used to express the meanings that are expressed using two words, "tree" and "wood" in English. Overall, these observations suggest that natural languages have a very specific property: some words have more meanings than others.

The unequal distribution of meanings across words is predicted by the simplicity-informativeness trade-off hypothesis (Zipf, 1945; Kemp & Regier, 2012; Kirby, Tamariz, Cornish, & Smith, 2015; Carr, Smith, Culbertson, & Kirby, 2020). This hypothesis states that the structure of the lexicon reflects a trade-off between simplicity – the number of words that speakers need to remember and produce in communication, and informativeness – the effort required by hearers to identify the correct reference of a word. Under this trade-off, languages are believed to find an optimal configuration between two extreme states; unification and diversification (Zipf, 1945). Overall, this approach correctly predicts the existence of ambiguity in different category systems ranging from numerical systems (Xu, Liu, & Regier, 2020) to color terms (Regier, Kay, & Khetarpal, 2007), and these results can be expanded to lexica in general (Kemp, Xu, & Regier, 2018).

Although the simplicity-informativeness trade-off does predict ambiguity in the lexicon, it does not predict which wordforms are more susceptible to acquiring more meanings. Such statistical regularities between the properties of words/wordforms and the number of meanings associated with them have been evidenced cross-linguistically. For instance, George Kingsley Zipf has noted that words that have more meanings tend to be more frequent; this regularity is known as Zipf's meaning-frequency correlation (Zipf, 1945; Casas, Hernández-Fernández, Catala, Ferrer-i Cancho, & Baixeries, 2019). To our knowledge, the mean-frequency correlation has only been evidenced in 10 different languages (Bond, Janz, Maziarz, & Rudnicka, 2019; Català, Baixeries, Ferrer-i Cancho, Padró, & Hernández-Fernández, 2021) (see also (Piantadosi et al., 2012)).

To explain Zipf's meaning-frequency correlation, as

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

well as provide a general theory of linguistic ambiguity, (Piantadosi et al., 2012) proposed a hypothesis, stating that language users benefit from ambiguity. In this paper, the authors argue that lexical ambiguity might be an efficient feature of a communicative system when the context is informative about what is being communicated. Ambiguity would thus allow the reuse of word forms that are the most easily produced, i.e. words that are shorter, more frequent and more phonotactically probable. They proposed that the reuse hypothesis might explain the uneven distribution of the number of meanings across words in different languages. To test this idea, (Piantadosi et al., 2012) used data from English, French and Dutch to show that the most ambiguous words in the lexicon are frequent, short, and have a high phonotactic probability. Overall, their results suggest that words are selected for reuse on the basis of communicative efficiency. They interpreted this result as direct evidence of ambiguity being a desirable property of language (Gibson et al., 2019), and not a property that is detrimental to communication, *contra* (Chomsky, 2002, p. 107).

The interpretation of this result has recently been challenged by computational modelling (Caplan et al., 2020; Trott & Bergen, 2020). In these studies, the authors attempted to construct null models devoid of selection for efficient communication. These models relied on generating lexica using n-gram models and randomly allocating meanings to each wordform in the simulated lexica. Both of these models yielded similar correlations to the ones found in (Piantadosi et al., 2012), such as the negative correlation between number of meanings and word lengths. The authors of both studies argued that these correlations can thus be explained by purely random allocation of meanings to words, as implemented in their respective models. However, (Pimentel, Teufel, Mahowald, & Cotterell, to appear) highlighted that those approaches are invalid, as they rely on n-gram models as a baseline, which do not reproduce many of the properties of the natural lexica, such as word rank-frequency distributions. To support their claim, Pimentel et al. replicated the results from (Caplan et al., 2020) using a Long short-term memory neural network, which has been shown to provide much more accurate results in word-generation tasks (Pimentel, Roark, & Cotterell, 2020). Under these conditions, this null model was not able to replicate the correlations from the (Piantadosi et al., 2012) study. Therefore, the null models present in the existing literature do not provide good evidence against the hypothesis stating that ambiguity is a result of selection for efficiency. Overall, this suggests that there remains a big explanatory gap that is still not filled by the existing literature.

Altogether, these results call for a refinement of the current theories of ambiguity. Previous approaches seem to be incomplete, as they largely ignore several key properties related to ambiguity. First, the existing literature seems to focus heavily on homophony, whereby words label unrelated meanings (the English words such as "tail" and "tale", which have the same phonological form /'teɪl/), largely ignoring or confounding it with polysemy, whereby words label multiple related meanings (the English word "run", which can mean both an activity (noun) and an action (verb), amongst other meanings). The only exception being (Piantadosi et al., 2012), where the authors analyzed data on polysemy from WordNet (Miller, 1995). By one estimate, in English, 7% of the words are homophones and 84% are polysemous (Rodd, Gaskell, & Marslen-Wilson, 2002). Overall, this suggests that polysemy should not be excluded when modelling ambiguity, as it seems to be much more prevalent than homophony. Second, understanding why ambiguity arises in the lexicon requires an understanding of how new meanings are introduced in the lexicon. Suppose that speakers need to communicate a novel meaning. There are several options available to them; they could reuse an existing word, thus leading to ambiguity or they could create a new word or borrow one from another language. Invention or borrowing are not negligible ways of introducing new meanings in the lexicon. To illustrate, (Tadmor, Haspelmath, & Taylor, 2010) explored a dataset containing terms for 1460 concepts collected for 41 languages from diverse regions of the world. In this rather small set of words, 25% of of words are borrowed when taking the average across languages. The borrowing rates are the highest across nouns: up to 40% of nouns were borrowed. Considering the rather small size of this dataset, this data indicates that borrowings represent a significant share of the lexicon. Additionally, speakers can proceed with coining entirely new words; for instance, words such as *whataboutism* and *quaranteen* were introduced in English during the last two years. These words correspond to two very specific meanings; for the former, a specific way of arguing, and, for the latter, young people spending their teenage years on quarantine during the COVID-19 pandemic. Overall, this suggests that reuse is not the only option when coining a new meaning, and events such as borrowings or word inventions should also be considered. However, it is also important to distinguish the acquisition of additional meanings, which is only possible with reuse, as inventions or borrowings usually introduce a new meaning together with a new word, and therefore no additional meanings are acquired. And, finally, the processes of semantic change, which include the introduction of new meanings to the lexicon, as well as meaning replacement, are diachronic in nature (Hamilton, Leskovec, & Jurafsky, 2016a). Therefore, any theory that aims at explaining regularities such as Zipf's meaning-frequency correlation should also be diachronic. This is further supported by findings linking word frequency to rates of language change (Pagel, Atkinson, & Meade, 2007; Pagel, Beaumont, Meade, Verkerk, & Calude, 2019) and to the probability of semantic shifts (Hamilton, Leskovec, & Jurafsky, 2018). Overall, this calls for a revision of the current approaches to ambiguity in favor of one that takes into account these properties to answer the question, Why do some words have more meanings than others?

A mechanism for acquiring new meanings

Here, we are proposing to employ the neutral model approach, which was introduced in (Neiman, 1995; Bentley, Hahn, & Shennan, 2004) for cultural evolution research. This approach consists of building models that assume no selection pressures when modelling cultural evolution events. Recently, these types of models have been adapted to investigate language change (Hamilton, Leskovec, & Jurafsky, 2016b; Newberry, Ahern, Clark, & Plotkin, 2017; Karsdorp, Manjavacas, Fonteyn, & Kestemont, 2020). Another benefit of adopting such an approach is that the assumption of the absence of selection can serve as a null model. This is very important for our study, since (Piantadosi et al., 2012) essentially proposed that words are *selected* for reuse based on their length and phonotactic complexity, as well as their frequency, a claim their critics dispute (Caplan et al., 2020; Trott & Bergen, 2020).

To test whether a selection pressure for communicative efficiency may explain patterns of ambiguity in natural language lexica, we propose an alternative mechanism for generating ambiguity that stems from a diachronic perspective. This mechanism is based on the following three basic assumptions. First, some words stay present in the lexicon for much longer periods of time than others. Indeed, it has been shown that during the process of language change, some words have much higher replacement rates than others (Pagel et al., 2007). Second, ambiguity arises in the lexicon because speakers reuse existing wordforms to denote new meanings (Piantadosi et al., 2012; Ramiro, Srinivasan, Malt, & Xu, 2018). Third, the possibility of reusing a wordform is conditional on the presence of that form in the lexicon at time t; trivially, if a wordform is not in the lexicon, it cannot be reused, as it is unavailable for speakers. Based on these assumptions, we expect "older" words to be reused much more often than "younger" ones. Therefore, we hypothesize that ambiguity should be better predicted by a word's longevity rather than by other characteristics such as length or frequency.

Here, we are proposing a neutral model inspired by the Full-sampling Neutral Model from (Ruck, Alexander Bentley, Acerbi, Garnett, & Hruschka, 2017) which shares its essential premises with the mechanism outlined above. Critically, this model is devoid of selection pressures for communicative efficiency, yet still allows ambiguity to develop. The process of reuse is truly independent of word characteristics such as frequency; reused words are randomly sampled from the list of unique words already present in the lexicon. Yet, because our model simulate lexica diachronically, reuse is directly affected by a word's longevity. With this model, we aim at showing that frequency-ambiguity correlations per se does not provide direct evidence for efficient communication pressures operating at the level of language. To be clear, if the model is able to reproduce the frequency-ambiguity correlation, this would suggest that no selection pressure is necessary for such meaning correlation to arise in the lexicon. This would not imply that longevity is the main factor influencing ambiguity in natural languages, but rather, that taking frequency-meaning correlations to exemplify the role of communicative efficiency in language is not warranted.

Model

This model has 6 parameters, which are summarized in Table 1. Given these parameters, the model works as follows. During the initial stage, the model starts by generating N types (unique words), which are represented as unique identifiers in the model. This is important, since the model presented here does not use natural language data to generate random words, contrary to previous approaches (Dautriche, Mahowald, Gibson, Christophe, & Piantadosi, 2017; Caplan et al., 2020; Trott & Bergen, 2020, 2022). Each type is taken to stand for a word that is associated with one or several meanings. Initially each type is associated with exactly one meaning. In the model, S corresponds to the number of unique meanings and N corresponds to the number of unique types with which the model is initialized. The parameters N and S are chosen such that $S > N$ to reflect the optimization problem of representing a large set of meanings with a limited amount of words. The model assigns the remaining S-N meanings to these N types by randomly sampling types S-N times and adding one to the meanings count for each sampled type. Finally, a set of p tokens is randomly sampled from the initial set of N types. Once the initial stage is completed, the following procedure is then repeated t times. The next set of tokens of the size p is sampled by either copying a token from the previously generated set with the probability $1-\mu$ or replacing it with a newly generated token with the probability μ. Then, at each time-step t the list of unique types is collected, and each type has a probability k of being reused (i.e. assigned an additional meaning). At each iteration, each type has a number of statistics that are stored. These include its longevity – the number of time-steps for which it is present in the population, its frequency – number of occurrences in the sample of the size p, its number of meanings and finally, its extinction – if it is removed at time-step t the value is set to 1.

Table 1: Description of the parameters of the model

Parameter	Description
N	Number of types
S	Number of senses
p	Number of tokens to sample
μ	Rate of new word appearance
k	Rate of reuse
t	Number of time-steps

This model has several important properties. The introduction of new types and corresponding meanings (rate of new word appearance μ) is reflecting the processes of borrowing and word invention in natural languages. The use of sampling and copying to generate p tokens at each generation models the drift-like process by which some types gain in popularity, and some types just disappear from the lexicon. A similar procedure was used in (Ruck et al., 2017), and their model yielded natural power-law like distribution of word frequencies as found in human languages.

Power law and the growth of the lexicon

First, we are demonstrating the validity of this model by showing that several properties of natural lexica are accurately simulated in our model. Let's consider a model with the following parameters; $S = 6,000$, $N = 3,042$, $\mu = 0.01$, $k = 0.02$, $p = 100,000$ and $t = 300$. Here, the choice of the number of meanings (S) and the number of types is motivated by the data extracted from English WordNet (Miller, 1995). We counted the number of unique words in the English WordNet (77636), and we added the respective number of synsets to get the total number of meanings (153270). Then, we took the ratio (0.507) and scaled the parameters in our model according to this ratio. The choice of μ and k is justified below. Here, we are only considering the data from the last time-step t_{300}. First, our model does reproduce the power law relationship between word rank and word frequencies, otherwise known as Zipf's law (Zipf, 2016 (1949); Piantadosi, 2014), as shown on Figure 1. The frequency-rank relation characteristic of natural languages is present in the model, which was expected since our implementation closely follows the model from (Ruck et al., 2017). In this paper, the authors reported that their model also yielded such power law-like distributions. On the contrary, previous null models of the lexica (Caplan et al., 2020; Trott & Bergen, 2020) do not seem to be capable of producing such frequency distributions.

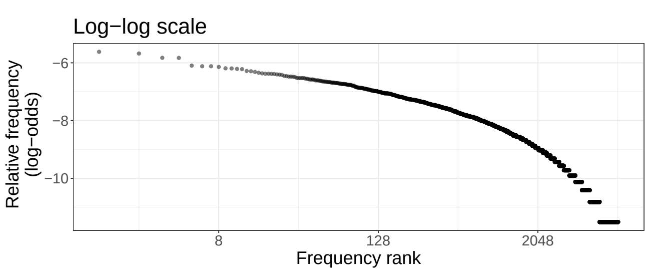

Figure 1: Word rank-frequency distribution in the last time-step of the model (t_{300}). Both relative frequency and rank are log-transformed.

Furthermore, the changes in the number of meanings and the number of unique types is examined below. As both the reuse and the addition of new types produce new meanings, the total number of meanings is expected to change during the running of the model. Same holds for the number of unique types, as borrowings increase the number of words, while some types can also be lost during copying. These processes are summarized on the figure below, where the number of unique types and the total amount of meanings are shown. Figure 2 shows that both the number of types and the number of meanings grow significantly until about t_{80}. After this, the number of of types stalls at around 7.500, while the number of meanings becomes stable at around 15.000. This can be contrasted with the model from (Ruck et al., 2017) and the model from (Caplan et al., 2020), where the number of types grows exponentially. Overall, the proposed model does accurately reproduce two very important properties of natural languages; the power law like frequency distribution of words, and the stable (at least: non-exponential) nature of the growth of the lexicon.

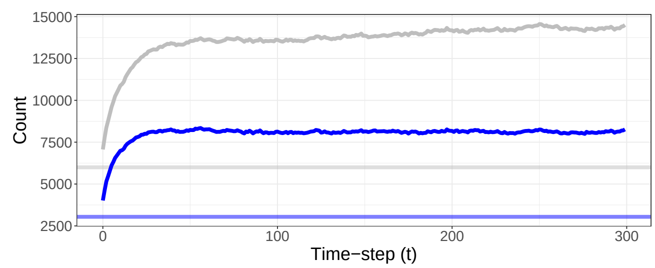

Figure 2: Total number of meanings (grey) and unique types (blue) during each time-step. Horizontal lines indicate the corresponding starting values (6000 and 2000, respectively).

Longevity and number of meanings correlation

Using the same model, we probed the correlation between longevity and number of meanings. As highlighted earlier, this is an essential property of the model, since it provides direct evidence for the mechanism proposed in this paper. The plot below shows the relation between longevity of a type – the number of iterations it was present in the model, and the number of meanings it has. This relation is contrasted with the frequency - number of meanings relation:

As shown on Figure 3, longevity is a much better predictor for the number of meanings a word has than its relative frequency. This holds true even when considering partial correlations: the meaning-longevity correlation when controlled for frequency (ρ = 0.5) is greater than the meaning-frequency correlation when controlling for longevity (ρ = -0.04).

Variation of the rate of new word appearance and reuse rate

The model described in this paper has 6 parameters, out of which the rate of borrowing μ and the rate of reuse k are the most important ones, as they are responsible for the relation between longevity (the number of iterations for which a type is present in the lexicon) and the number of meanings this type has. We expect that the correlation between longevity and the number of meanings is always greater than the corre-

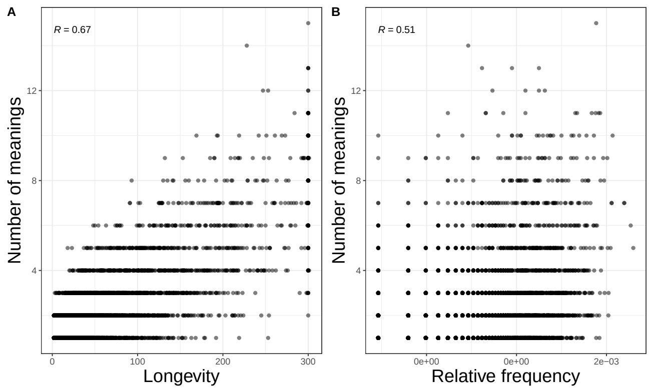

Figure 3: Correlation between longevity and number of meanings (A) and relative frequency (log-scale) and number of meanings (B) n the last time-step of the model (t_{300}). Each point corresponds to a unique word. The values on the top left corner indicate the correlation coefficients (Spearman's rho) and the corresponding p-values.

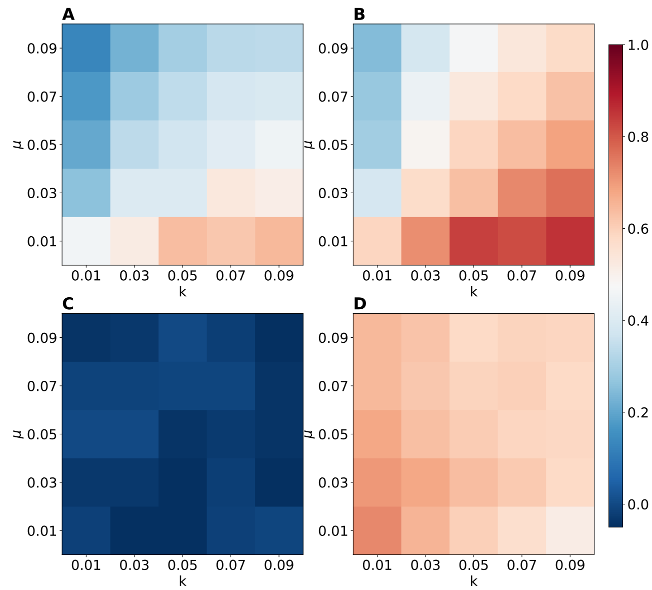

Figure 4: Results of the first simulation. (A) corresponds to frequency-meaning correlation (Spearman's rho), (B) longevity-meaning, (C) frequency-meaning (controlled for longevity) and (D) longevity-meaning (controlled for frequency). Warmer colours correspond to higher correlation coefficients, while colder colors correspond to lower correlation coefficients. The value of the reuse rate parameter k is plotted on the x-axis, while the values on the y-axis correspond to the replacement parameter μ.

lation between frequency and the number of meanings. Additionally, we expect that, when controlled for longevity, the correlation between frequency and the number of meanings should be smaller than the correlation between longevity and the number of meanings when controlled for frequency.

To test this prediction, we vary μ and k parametrically from 0.01 to 0.9 with a step of 0.2, resulting in 25 combinations of pairs of parameters. The other parameters are fixed: $N = 600$, $S = 305$, and $p = 10.000$. Essentially, they correspond to the parameters of the model used above, but they are scaled down by dividing the initial parameters by 10 for computing reasons. The model is then initialized with these fixed parameters except for μ and k, as their values are taken from the 25 combinations defined above. Then, for each μ and k combination, we compute the correlation (Spearman's rho) between longevity and number of meanings, the correlation between frequency and number of meanings, the partial correlation between longevity and number of meanings (controlling for frequency) and the partial correlation between frequency and number of meanings (controlling for longevity). The heatmap on Figure 4 shows the results of this simulation. As shown on the heatmap, the model behaves as expected. First, the frequency - number of meanings correlation is always equal to zero when controlled for longevity under any combination of μ and k. The longevity - number of meanings correlation is, on the other hand, always greater than 0.6 even when controlled for frequency. This suggests that frequency, in this model, is simply a confounding variable, and the longevity of a word is the only variable influencing the number of meanings it has in this model, as proposed in the section describing the proposed mechanism. Additionally, the results presented on Figure 4 highlight that this model is valid in the sense that there is significant variation when the parameters are being changed (panels A and B). However, when the respective correlations are controlled, the results are uniform (panels C and D), indicating that the predictions made in the section devoted to mechanism description are valid under any combination of the parameters.

Discussion

In this paper, we introduce a new model aimed at representing the causal mechanism behind the appearance of linguistic ambiguity. This paper follows the recent trend in the literature (Caplan et al., 2020; Trott & Bergen, 2020), which challenges the idea that frequency-meaning correlations are a marker of efficient communication (Piantadosi et al., 2012) by providing null models ostensibly lacking selection of forms based on their communicative efficiency. The model proposed in this paper has several improvements in comparison to previous approaches. First of all, this model is truly free of any selective pressures with regards to reuse.The earlier approaches used the n-gram models to generate random words (see (Dautriche et al., 2017) for the introduction of this methodology). Since these models are trained on data coming from natural languages, they are prone to copying various biases from natural languages. For instance, since n-gram models are prone to overfitting (Pimentel et al., 2020), they were amplifying the length-frequency bias. Therefore, the randomly generated lexica were skewed towards containing shorter words on average. Such that, even if words were randomly chosen for reuse as in (Caplan et al., 2020), the length bias was still present in this procedure. In the present paper, the model does not interact with natural language data in any way. Additionally, words in this model are devoid from any features such as phonotactic probability or frequency, allowing for a truly neutral, selection free reuse procedure to take place. Secondly, the model reflects the diachronic nature of semantic change, which is essential when describing such phenomena as linguistic ambiguity.

The results of the simulations discussed above are indicating that the model presented here conforms to a mechanism where there is no direct causal effect of word frequency on the number of meanings. Instead, the number of meanings is directly influenced by the word's longevity – the longer a word is present in the lexicon, the higher its chance of being reused. Overall, this model reflects a process in which ambiguity emerges in the lexicon but is not motivated by communicative efficiency: the wordforms that are ambiguous are not selected for reuse based on their properties such as frequency, length, or phonotactic probability. The data obtained from this model can be used as a null assumption, serving to determine whether the process of reuse of wordforms is selection-free or not.

However, our model is not devoid of limitations. One may worry that the longevity-meanings correlation is in a way "built into" the model. This is not entirely true, since there are possible parameter combinations that will produce no longevity-meaning correlations. For instance, if the rate of reuse (k) is quite low, while the rate of replacement (μ) is high, the longevity-frequency correlation will be essentially cancelled out, since the words would be replaced faster that they would gain new meanings. As an example, consider $\mu = 0.09$ and $k = 0.01$ on Figure 4 (panel B); the longevity-meaning correlation is very weak under this combination of parameters. Which means that the meaning-frequency correlation is not an inherent property of this model, but rather the result of the specific mechanism embedded into it. Another limitation of this paper in particular is that we did not contrast our findings with real word data. However, this issue can be resolved in future work, since the data generated by this model can be compared with real data to determine whether those pressures are shaping the lexicon or not. The data for comparison can be a combination of word longevity approximations from large diachronic corpora combined with the number of senses estimated from lexical databases such as WordNet (Miller, 1995). Those kinds of databases have been extensively used in previous studies focusing on diachronic semantic change, such as (Karjus, Blythe, Kirby, & Smith, 2020; Ramiro et al., 2018). The model parameters can be matched to those of real data. For instance, the number of senses and the number of unique words (types) can be collected from the real data and plugged into the model, together with the time-step parameter, which will be matching the historical scale of the chosen corpus. Using the data generated with a matching model, we can compare real statistics, such as the correlation between longevity and the number of meanings, as presented in this paper. Additionally, we have not considered other relevant phenomena such as word lengths and phonotactic complexity, which were discussed both in (Piantadosi et al., 2012; Caplan et al., 2020). However, the presented model allows to account for those variables as well. For instance, each type in the model can receive a length which would be drawn from a discrete distribution that resembles the lengths distribution found in a particular natural language. Then, the simulated correlations could be again compared with natural language data.

Acknowledgments

The authors would like to thank the members of the Mint team for their useful comments, as well as the anonymous reviewers for their comments and suggestions.

Data availability

The repository with data and code needed to reproduce the analysis presented in this paper can be found here: `https://osf.io/sh7zy/?view_only=a459b10a37f842d0b4762acd8de22619`

References

Bentley, R. A., Hahn, M. W., & Shennan, S. J. (2004). Random drift and culture change. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1547), 1443–1450.

Bond, F., Janz, A., Maziarz, M., & Rudnicka, E. (2019). Testing zipf's meaning-frequency law with wordnets as sense inventories. In *Proceedings of the 10th global wordnet conference* (pp. 342–352).

Caplan, S., Kodner, J., & Yang, C. (2020, December). Miller's monkey updated: Communicative efficiency and the statistics of words in natural language. *Cognition*, *205*, 104466. doi: 10.1016/j.cognition.2020.104466

Carr, J. W., Smith, K., Culbertson, J., & Kirby, S. (2020, September). Simplicity and informativeness in semantic category systems. *Cognition*, *202*, 104289. doi: 10.1016/j.cognition.2020.104289

Casas, B., Hernández-Fernández, A., Catala, N., Ferrer-i Cancho, R., & Baixeries, J. (2019). Polysemy and brevity versus frequency in language. *Computer Speech & Language*, *58*, 19–50.

Català, N., Baixeries, J., Ferrer-i Cancho, R., Padró, L., & Hernández-Fernández, A. (2021). Zipf's laws of meaning in catalan. *PloS one*, *16*(12), e0260849.

Chomsky, N. (2002). On Nature and Language.

Dautriche, I., Mahowald, K., Gibson, E., Christophe, A., & Piantadosi, S. T. (2017, June). Words cluster phonetically beyond phonotactic regularities. *Cognition*, *163*, 128–145. doi: 10.1016/j.cognition.2017.02.001

François, A. (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In M. Vanhove (Ed.), *Studies in Language Companion Series* (Vol. 106, pp. 163–215). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/slcs.106.09fra

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019, May). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, *23*(5), 389–407. doi: 10.1016/j.tics.2019.02.003

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016a). Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing* (Vol. 2016, p. 2116).

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016b). Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 2116–2121). Austin, Texas: Association for Computational Linguistics. doi: 10.18653/v1/D16-1229

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2018, October). *Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change* (No. arXiv:1605.09096). arXiv. (Comment: Association for Computational Linguistics (ACL), 2016. Minor corrections; improved methodology for Section 3)

Karjus, A., Blythe, R. A., Kirby, S., & Smith, K. (2020, May). Challenges in detecting evolutionary forces in language change using diachronic corpora. *Glossa: a journal of general linguistics*, *5*(1). doi: 10.5334/gjgl.909

Karsdorp, F., Manjavacas, E., Fonteyn, L., & Kestemont, M. (2020). Classifying evolutionary forces in language change using neural networks. *Evolut. Hum. Sci.*, *2*, e50. doi: 10.1017/ehs.2020.52

Kemp, C., & Regier, T. (2012, May). Kinship Categories Across Languages Reflect General Communicative Principles. *Science*, *336*(6084), 1049–1054. doi: 10.1126/science.1218811

Kemp, C., Xu, Y., & Regier, T. (2018, January). Semantic Typology and Efficient Communication. *Annu. Rev. Linguist.*, *4*(1), 109–128. doi: 10.1146/annurev-linguistics-011817-045406

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015, August). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102. doi: 10.1016/j.cognition.2015.03.016

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*(11), 39–41.

Neiman, F. D. (1995). Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in illinois woodland ceramic assemblages. *American Antiquity*, *60*(1), 7–36.

Newberry, M. G., Ahern, C. A., Clark, R., & Plotkin, J. B. (2017, November). Detecting evolutionary forces in language change. *Nature*, *551*(7679), 223–226. doi: 10.1038/nature24455

Pagel, M., Atkinson, Q. D., & Meade, A. (2007, October). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*(7163), 717–720. doi: 10.1038/nature06176

Pagel, M., Beaumont, M., Meade, A., Verkerk, A., & Calude, A. (2019, April). Dominant words rise to the top by positive frequency-dependent selection. *Proc. Natl. Acad. Sci. U.S.A.*, *116*(15), 7397–7402. doi: 10.1073/pnas.1816994116

Piantadosi, S. T. (2014, October). Zipf's word frequency law in natural language: A critical review and future directions. *Psychon Bull Rev*, *21*(5), 1112–1130. doi: 10.3758/s13423-014-0585-6

Piantadosi, S. T., Tily, H., & Gibson, E. (2012, March). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280–291. doi: 10.1016/j.cognition.2011.10.004

Pimentel, T., Roark, B., & Cotterell, R. (2020, December). Phonotactic Complexity and Its Trade-offs. *Transactions of the Association for Computational Linguistics*, *8*, 1–18. doi: 10.1162/tacl_a_00296

Pimentel, T., Teufel, S., Mahowald, K., & Cotterell, R. (to appear). Random typing models are not as efficient as natural languages. , 17.

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. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.1714730115` doi: 10.1073/pnas.1714730115

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*(4), 1436–1441.

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.

Ruck, D., Alexander Bentley, R., Acerbi, A., Garnett, P., & Hruschka, D. J. (2017, September). Role of Neutral Evolution In Word Turnover During Centuries of English Word Popularrity. *Advs. Complex Syst.*, *20*(06n07), 1750012. doi: 10.1142/S0219525917500126

Tadmor, U., Haspelmath, M., & Taylor, B. (2010, October). Borrowability and the notion of basic vocabulary. *DIA*, *27*(2), 226–246. doi: 10.1075/dia.27.2.04tad

Trott, S., & Bergen, B. (2020, December). Why do human languages have homophones? *Cognition*, *205*, 104449. doi: 10.1016/j.cognition.2020.104449

Trott, S., & Bergen, B. (2022, August). Languages are efficient, but for whom? *Cognition*, *225*, 105094. doi: 10.1016/j.cognition.2022.105094

Xu, Y., Liu, E., & Regier, T. (2020, 08). Numeral Systems Across Languages Support Efficient Communication:

From Approximate Numerosity to Recursion. *Open Mind*, *4*, 57-70. Retrieved from https://doi.org/10.1162/opmi_a_00034 doi: 10.1162/opmi_a_00034
Zipf, G. K. (1945, October). The Meaning-Frequency Relationship of Words. *The Journal of General Psychology*, *33*(2), 251–256. doi: 10.1080/00221309.1945.10544509
Zipf, G. K. (2016 (1949)). *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio Books.