Predicting strategy choice in word formation: A case study of reuse and compounding

Aotao Xu[1,2] **(a26xu@cs.toronto.edu)**
Charles Kemp[3] **(c.kemp@unimelb.edu.au)**
Lea Frermann[2] **(lea.frermann@unimelb.edu.au)**
Yang Xu[1,4] **(yangxu@cs.toronto.edu)**

[1]Department of Computer Science, University of Toronto
[2]School of Computing and Information Systems, University of Melbourne
[3]School of Psychological Sciences, University of Melbourne
[4]Cognitive Science Program, University of Toronto

Abstract

Natural language expresses new concepts by reusing existing words or coining new ones. Previous studies have examined these word formation strategies separately through a functional lens, but it is unclear why one strategy might be preferred over another. In this study, we hypothesize that communicative and cognitive efficiency might predict the choice between lexical reuse and compounding for expressing an emerging concept. We test our hypothesis by developing a computational analysis of English word meanings that emerged over the past century. Our results suggest that strategy choice may be explained partly by a pressure for least effort. Our work contributes a novel connection between strategy choice in word formation and functional theories of language.

Keywords: the lexicon; word formation; word meaning extension; compounding; efficiency

Introduction

Natural language adapts to an evolving culture by assigning word forms to emerging concepts. For example, the English lexicon used the form *car* to express a version of the motor vehicle as this concept emerged during the early twentieth century. This example reflects two common strategies through which the lexicon expresses emerging meanings: reusing an existing form, such as the horse-drawn *car*, or combining existing forms into a compound such as *motor vehicle*. What factors predict the word formation strategy for expressing an emerging concept? Here we address this question with a computational analysis of historical cases of word reuse and compounding in English.

We begin our investigation from a functional perspective suggesting that language reflects communicative and cognitive efficiency (e.g., Rosch, 1978; Zipf, 1949). Existing work in linguistics has extended this perspective to account for word formation (Dressler, 2005; Štekauer, 2005), the study of patterns with which the lexicon forms new words (Marchand, 1960). This functional theory of word formation is supported by empirical work on historically attested cases of word meaning extension (Y. Xu, Malt, & Srinivasan, 2017; Ramiro, Srinivasan, Malt, & Xu, 2018), conventionalized complex words (A. Xu, Kemp, Frermann, & Xu, 2022), and loan words (Monaghan & Roberts, 2019), but these different word formation strategies have typically been examined in isolation. We extend these previous studies by suggesting that communicative and cognitive efficiency might also predict the specific strategy for expressing an emerging meaning, and in this initial work we focus on the strategy choice between reuse and compounding.

One established account of communicative efficiency originates from the observation that word frequency and word length tend to be anti-correlated (Zipf, 1949). This anti-correlation reflects an optimization of the average word length in the lexicon, hence allowing meanings frequently talked about to be expressed in shorter forms and those rarely talked about to be expressed in longer forms. Recent work has formalized this idea in information-theoretic terms and rigorously examined this tendency in attested lexicons (Ferrer-i Cancho, Bentz, & Seguin, 2022; Mollica et al., 2021; Pimentel, Nikkarinen, Mahowald, Cotterell, & Blasi, 2021). This least effort account has also been supported by experiments on artificial language learning (Kanwal, Smith, Culbertson, & Kirby, 2017) and repeated reference games (Krauss & Weinheimer, 1964; Hawkins et al., 2022). Here we hypothesize that the lexicon adapts to the *communicative need* of emerging meanings, i.e., the frequency with which the concept is encountered: as meanings enter the lexicon, their need should constrain the length of their corresponding word forms, so that the lexicon remains relatively compact in a Zipfian sense. In particular, since the plausible compounds for expressing a given meaning tend to be longer than plausible existing word forms (e.g., *motor car* is longer than *car*), the view of least effort predicts that the lexicon should prefer expressing high-need emerging meanings with word reuse instead of compounding.

We also consider an alternative account of efficiency rooted in earlier theories from cognitive linguistics (Lakoff, 1987; Geeraerts, 1997). Using formal categorization models and large-scale historical data, this line of work found that the lexicon tends to express novel meanings with existing words that are semantically similar (Ramiro et al., 2018; Grewal & Xu, 2021; Yu & Xu, 2021), which may reflect ease in learning (Srinivasan, Al-Mughairy, Foushee, & Barner, 2017; Floyd & Goldberg, 2021). However, there might not always be an existing word available that is sufficiently similar to a novel meaning. Certain concepts (e.g., *radioactivity* or *nuclear winter*) may be highly novel relative to the concepts ex-

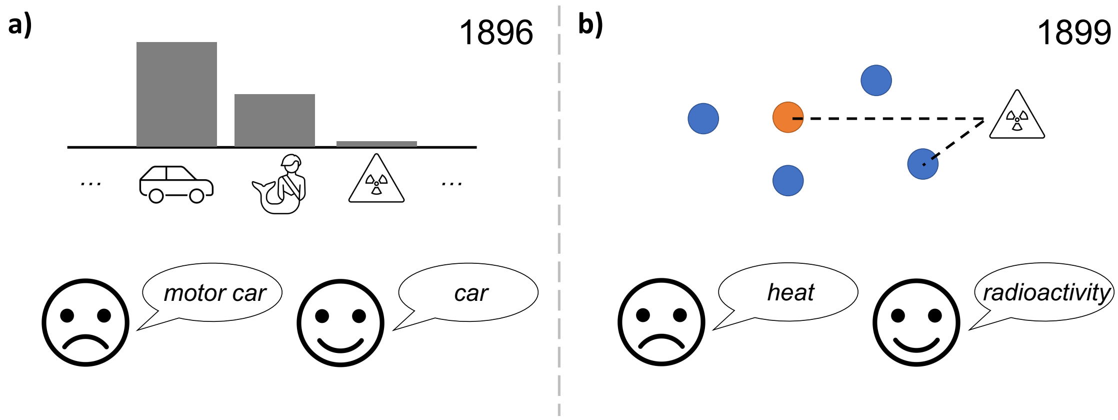

Figure 1: An illustration of our a) least effort hypothesis and b) novelty hypothesis. In panel a), grey bars represent the distribution of communicative need for different concepts, and the speaker prefers to reuse the form "car" for a meaning with high communicative need. In panel b), each dot or symbol corresponds to a meaning in similarity space; red and blue dots correspond to existing meanings or meanings of potential compounds, respectively. Here the speaker prefers a compound form for a meaning that is far from the meanings expressed by existing words. In both panels, the upper right corner shows the year of first occurrence for emerging concepts according to the Oxford English Dictionary (OED).

pressed by the existing lexicon, thereby impeding the learning or communication of those emerging concepts. We therefore hypothesize that the lexicon should avoid reusing existing words to express highly novel meanings. Intuitively, compounding might be a more efficient choice in the case of high novelty, since a large number of conceptual combinations covers novel meanings more compactly in similarity space than existing meanings.

We illustrate our hypotheses in Figure 1. In the left panel, the least effort hypothesis postulates that high-need meanings are more likely to be expressed by reuse than by compounding. In the right panel, the novelty hypothesis postulates that high-novelty meanings should prompt compounding over reuse. In the following, we introduce our dataset of (historical) word meanings and our operationalization of need and novelty. We then evaluate the least effort and novelty hypotheses separately, and conduct a predictive analysis to test whether the two constraints can jointly predict the attested word formation strategy of meanings that emerged over the historical development of English.

Data

To investigate strategy choice in word formation, we used three sources of data. First, we used a large historical text corpus to measure communicative need and recreate historically existing lexicons. Second, we used an English dictionary that records words and their meanings to measure relative novelty. To analyze historically emerging meanings, we timestamped the first occurrence for a subset of the compounds and polysemous words.

Historical text data. We collected historical word usages from the Corpus of Historical American English (COHA; Davies, 2002). The corpus contains English text published between 1810 to 2009, spanning across four genres. To more accurately estimate the frequencies of historically rare words, we supplemented COHA with unigram frequencies provided by Google Books Ngrams (English 2012 version; Michel et al., 2011). In total, we obtained 605K lemma types from COHA and 465B word tokens between 1810 and 2012 from Google Books.

Dictionary data. We primarily used the record of established words and their meanings provided by Hu, Li, and Liang (2019), which was originally based on the English version of the Oxford Dictionary (OD)[1]. This dataset consists of $3,220$ frequent polysemous words, containing sense definitions and historical sense frequencies for every decade between 1810 and 2009, which are estimated using COHA and supervised word sense disambiguation[2]. Since their dataset only contains polysemous words, we supplemented this dataset with monosemous words from two sources: 1) we obtained $3,353$ compounds that appear in the Large Database of English Compounds (LADEC; Gagné, Spalding, & Schmidtke, 2019) and the online version of OD[3], and 2) we obtained $43,482$ COHA lemmas and their definitions that appear in archived webpages of OD[4]. In total, this provided us with a set of $77,359$ senses for $30,760$ words. Example sense definition are shown in Table 4.

Word sense emergence. Since OD was created for contemporary English, we manually timestamped the first occurrence of a subset of OD word senses that emerged in the 20th century.[5] We identified this subset by using a method of shortlisting sense definitions that contain cultural key-

[1] Note that this is different than the OED.
[2] This tags each polysemous word with one of its senses.
[3] `https://www.lexico.com/`
[4] `https://archive.org`
[5] Data can be found at `https://github.com/johnaot/strategy-prediction-data`

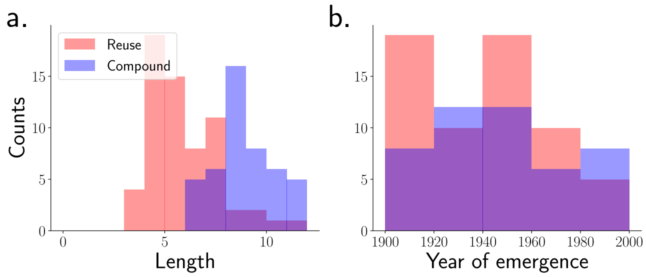

Figure 2: Descriptive statistics for our dataset of 20th-century word senses expressed by reuse and compounds.

aircraft, airport, broadcast, broadcasting, car, cinema, computer, data, electric, electrical, electronic, film, jazz, motor, phone, pilot, radio, software, television, video

Table 1: Set of cultural keywords.

words (Cook, Lau, McCarthy, & Baldwin, 2014). To identify these keywords, the first author manually selected culturally salient words that have changed the most in frequency since 1900 in COHA. Our keywords are shown in Table 1. We used these keywords to shortlist senses of polysemeous words from Hu et al. (2019) and compounds from LADEC. Since most compound words are nouns, we focused on noun senses in the shortlist. After timestamping the shortlist with the OED, we obtained 67 cases of reuse and 46 compounds, as well as their exact year of first occurrence in the OED. To guarantee positive communicative need[6], we filtered out senses that have zero frequency for two decades after their emergence. This left us with our final dataset of 63 meanings with an attested reuse form, and 46 meanings expressed with compounds.

Figure 2 shows descriptive statistics for our dataset. The senses we collected emerged during the whole span of the 20th century. The orthographic (word) lengths of senses expressed via reuse tend to be shorter than the lengths of compound words in the dataset.

Computational methods

To evaluate our hypotheses on word formation strategies for emerging meanings, we first define the space of meanings obtained from our datasets. We then define the measures we use to quantify the properties of emerging meanings.

Meaning space. Let $M = \{m_1, m_2, ..., m_k\}$ be the set of meanings we will use in our analyses. We defined each meaning m_i as a word sense we obtained from OD. We represented each meaning as a 768-d vector by embedding its sense definition with Sentence-BERT (Reimers & Gurevych, 2019)[7].

[6]The purpose of this consideration is visualization, but regardless our results do not change significantly.

[7]The model was trained on a corpus of contemporary texts, but we make the simplifying assumption that word senses have the same meanings over time.

We ensured that the embedding space reflects genuine semantic relatedness using three datasets of pairwise word similarity ratings: WordSim-353 (Finkelstein et al., 2001), SimLex-999 (Hill, Reichart, & Korhonen, 2015), and MEN (Bruni, Tran, & Baroni, 2014). We compared our sense embeddings based on sentence-BERT to a version where we represent senses by averaging the Word2Vec (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Grave, Bojanowski, Puhrsch, & Joulin, 2018) vectors of non-stopwords in their definitions. We also include a word (rather than sense) level method using pre-trained Word2Vec embeddings. This method solely serves an upper bound in the context of our word similarity validation task, but cannot be applied in our main analyses which rely on sense-level representations.

To evaluate these embeddings, we computed the spearman correlation of human ratings and the cosine distances between word embeddings. For sense-level embeddings, we represented each word by averaging the embeddings of its senses. The evaluation results are summarized in Table 2. We see that Sentence-BERT embeddings are significantly better than the OD (Word2Vec) baseline, possibly because the former is able to better capture the dependencies among words within each definition. While OD senses do not perform as well as the word-level Word2Vec upper bound, both correlate consistently and significantly with human similarity ratings. We proceeded with the Sentence-BERT representation, and leave its improvement for future work.

Communicative need of an emerging meaning. The need probability of a newly emerged meaning is difficult to measure, and in our historical framework we approximate it through the observed frequency of a meaning after emergence in a diachronic corpus. Suppose a meaning m_i emerged in year t, and suppose w is its attested word form. Let $p_x(w)$ be the relative frequency of w in year x according to Google Books; let $p_x(m|w)$ be the proportion of tokens of w expressing meaning m in year x, obtained from Hu et al. (2019). Since the frequency of new words tends to be sparse when they just emerged, we estimated the communicative need of m_i in year t, denoted $f(m_i)$, by averaging its frequency over a specified time window X:

$$f(m) = \frac{1}{|X|} \sum_{x \in X} p_x(m|w) p_x(w) \tag{1}$$

For robustness, we used two time windows: a historical window, where $X = \{t, t+1, ..., t+19\}$ for each m_i; and a contemporary window, where $X = \{2000, 2001, ..., 2012\}$. In our analyses, we multiplied the estimated need by a constant 10^6 to avoid numerical issues.

Relative novelty to the existing lexicon. To quantify the novelty of an emerging meaning m_i relative to the existing lexicon $\mathbb{L}_t$, we first defined $\mathbb{L}_t$. We started by dividing up the time period between 1880 and 2000 into consecutive intervals of 20 years, denoted $I_1, ..., I_6$. We then identified the list of senses that existed in I_{i-1} if year $t \in I_i$. Specifically, we automatically selected every sense m such that 1) if its attested

Dataset	OD (Sentence-BERT)	OD (Word2Vec)	Word2Vec	Sample size (n)
WordSim-353 (sim)	-0.576 ***	-0.077	-0.836 ***	181
WordSim-353 (rel)	-0.276 ***	-0.072	-0.731 ***	220
SimLex-999	-0.289 ***	0.111 **	-0.404 ***	964
MEN	-0.494 ***	0.007	-0.837 ***	2659

Table 2: Evaluation of embedding space using word similarity ratings. Each cell shows the spearman correlation between human ratings and embedding cosine distances, and embeddings with more negative correlations are better at capturing human ratings. The labels "**" and "***" denote significance at $p < 0.01$ and $p < 0.001$, respectively.

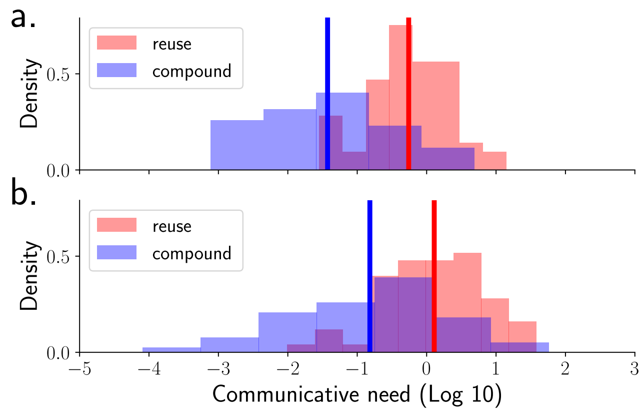

Figure 3: Communicative need estimated using a) historical data and b) contemporary data. Each vertical line indicates the mean over attested cases of reuse (red) or compounds (blue).

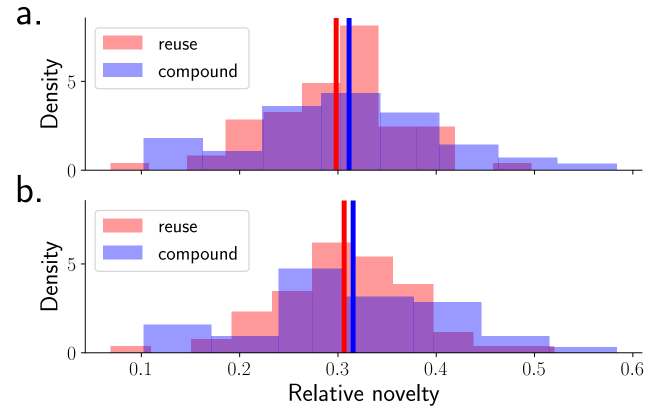

Figure 4: Relative novelty calculated using a) the nearest existing sense and b) the nearest existing word. Each vertical line indicates the mean over attested cases of reuse (red) or compounds (blue).

word w is polysemous, it satisfies $p_x(m|w) \geq 0.1$ for some $x \in I_{i-1}$, or 2) otherwise, we included it if its corresponding word w occurs at least 10 times during I_{i-1}. We defined $\mathbb{L}_t^{\text{sense}}$ as this list of existing senses. Additionally, we considered a word-level version where $\mathbb{L}_t^{\text{word}}$ is a list of word meanings, in which the meaning of word w is the average of its existing senses.

We measured the novelty of m_i relative to $\mathbb{L}_t$ as the cosine distance from m_i to $\mathbb{L}_t$:

$$d(m_i, \mathbb{L}_t) = \min_{m \in \mathbb{L}_t} \text{cos-dist}(m_i, m) \quad (2)$$

Intuitively, this measure characterizes the upper bound on the semantic similarity between m_i and the existing lexicon. A low novelty score (i.e., small cosine distance to the nearest existing word) implies the opportunity to achieve more cognitive efficiency via reuse, and vice versa[8].

Results

In this section, we first show our results for each of the two hypotheses we outlined previously. We then use the predictors motivated by these hypotheses jointly to predict the historically attested choices of word formation strategy, and qualitatively analyze our predictions.

[8]This is graphically illustrated in Figure 1b).

Evaluating the least effort hypothesis

Since the least effort hypothesis implies the lexicon prefers to express high-need meanings with reuse, we expected this preference to distinguish the average communicative need of reuse meanings and compound meanings. The comparison is illustrated in Figure 3. In both settings, the communicative need of meanings expressed by reuse tends to be higher than meanings expressed as compounds. To assess the statistical significance of this tendency, we compared the mean need of each group using independent t-tests. For needs estimated from historical data, we obtained $t(107) = 7.957$, $p < 0.001$. For needs estimated from contemporary data, we obtained $t(107) = 4.920$, $p < 0.001$. This provides evidence for our least effort hypothesis.

Evaluating the novelty hypothesis

Similar to the previous hypothesis, we compared the average novelty of reuse and compound meanings to assess the novelty hypothesis. The results are illustrated in Figure 4. We observe that on average, the distances for meanings expressed by compound tend to be higher than meanings expressed by reuse. Similar to the previous hypothesis, we compared the mean novelty of each group using t-tests. For distances computed at the sense level, we obtained $t(107) = -0.801$, $p = 0.425$. For distances computed at the word level, we ob-

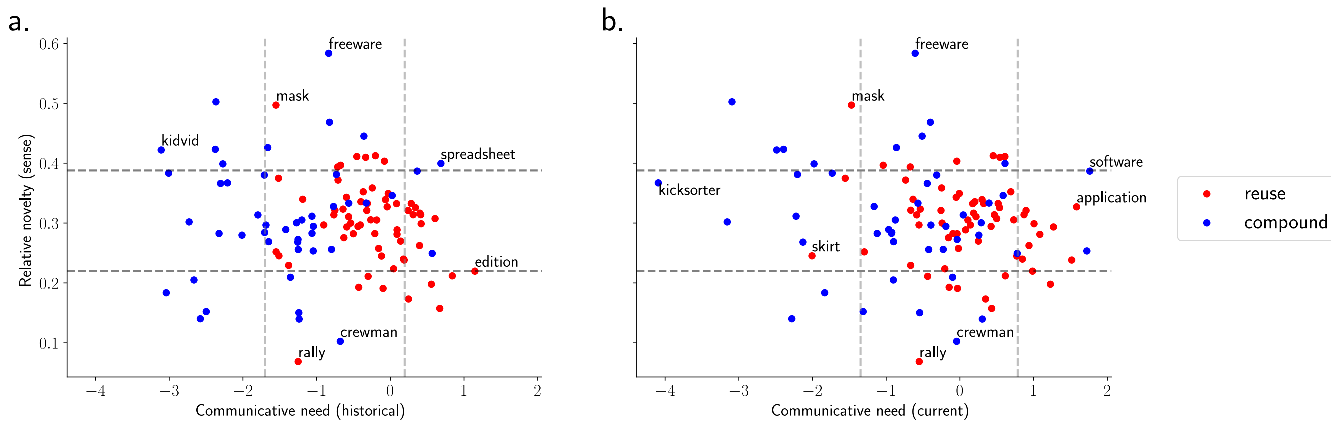

Figure 5: Scatter plots showing the communicative need and novelty of each emerging meaning. Each dot corresponds to an emerging meaning, and the extreme cases for each strategy are annotated. Dashed lines correspond to one standard deviation away from the mean over all meanings.

tained $t(107) = -0.538$, $p = 0.592$. In both cases, we do not find evidence for our novelty hypothesis.

Predicting strategy choice from need and novelty

After testing our hypotheses individually, we explored whether communicative need and novelty can jointly predict the attested strategy of an emerging meaning.

Since we focused on reuse and compounding, we formulate this task as a binary prediction problem. We proceeded by using a logistic regression model of the following form:

$$y(m) \sim \Big(1 + \exp(-(\beta_0 + \beta_1 f_{hist}(m) + \beta_2 f_{now}(m) + \beta_3 d(m, \mathbb{L}_t^{\text{sense}}) + \beta_4 d(m, \mathbb{L}_t^{\text{word}})))\Big)^{-1} \quad (3)$$

Here $y(m)$ refers to whether m is expressed by reuse or compound, and f_{hist}, f_{now} refer to the communicative need of m estimated using historical and contemporary frequencies. We implemented the model using `statsmodel` (Seabold & Perktold, 2010).

To evaluate the model, we trained it on word meanings that emerged before 1950 ($n = 66$), and we tested it using word meanings that emerged after 1950 ($n = 43$). The accuracy of the model is 0.767, which is higher than the percentage of the most frequent class, 0.558. We also tested the statistical significance of each individual predictor using a Wald test. The test results are summarized in Table 3. We observe that $f_{hist}(m)$ has the strongest effect size, while other predictors have no statistical significance.

To better understand how the predictors relate to reuse and compounding, we plotted individual meanings in Figure 5; meanings for labelled cases are detailed in Table 4. Since the novelty measures are highly colinear (Pearson correlation 0.978, $p < 0.001$, $n = 109$), we focus on plotting meanings with respect to sense-level novelty and both of the need measures. In both plots of Figure 5, we observe that meanings tend to have a wide range of novelty and communicative

Predictor	β	Z statistic	p-value
Intercept	−1.5353	−0.977	0.328
$f_{hist}(m)$	−2.8210	−3.488	< 0.001
$f_{now}(m)$	0.3194	0.618	0.537
$d_s(m, \mathbb{L}_t)$	3.7112	0.144	0.886
$d_w(m, \mathbb{L}_t)$	−7.6745	−0.309	0.757

Table 3: Summary of logistic regression results.

need. Aligned with our predictions, most meanings with a need higher than one standard deviation away from the mean tend to be expressed as reuse, whereas in the other direction most meanings tend to be expressed as compounds. In contrast, both low novelty and high novelty meanings are found across both groups.

We also observe in Figure 5a) that $f_{hist}(m)$ more strongly separates the two attested strategies than $f_{now}(m)$ in Figure 5b). This can be seen from the meanings with the most extreme communicative needs across the two strategies: for $f_{hist}(m)$, both the most needed and least needed reuse meanings (*mask* and *edition*) had higher need than their compound counterparts (*spreadsheet* and *kidvid*); however, for $f_{now}(m)$, the most needed in the compound group (*software*) actually has higher need than its counterpart in the reuse group (*application*). This suggests communicative need and strategy choice are sensitive to cultural changes over time.

Discussion

In this paper, we sought to explain the strategy choice between reusing an existing word and coining a novel form. Building on previous functional accounts of the lexicon (e.g., Ramiro et al., 2018; Mollica et al., 2021), we hypothesized that pressures for communicative and cognitive efficiency account for the choice between reuse and compounding. Specifically, our least effort hypothesis predicts that new meanings with high communicative need should be expressed by

Word	Time	Sense Definition
rally	1911	A long-distance race for motor vehicles over public roads or rough terrain, typically in several stages.
skirt	1912	A surface that conceals or protects the wheels or underside of a vehicle or aircraft.
edition	1934	A particular instance of a regular radio or television programme.
crewman	1937	A member of a group of people who work on and operate a ship, aircraft, etc., particularly one who is not an officer.
kicksorter	1947	A device for analysing electrical pulses according to amplitude.
kidvid	1955	Children's television or video entertainment.
mask	1956	A patterned metal film used in the manufacture of microcircuits to allow selective modification of the underlying material.
software	1958	The programs and other operating information used by a computer.
application	1959	A program or piece of software designed to fulfil a particular purpose.
freeware	1982	Software that is available free of charge.
spreadsheet	1983	An electronic document in which data is arranged in the rows and columns of a grid and can be manipulated and used in calculations.

Table 4: Examples of OD sense definitions and their year of emergence according to the OED.

reuse, whereas our novelty hypothesis predicts high-novelty meanings should be expressed by compounding. Using two operationalizations of communicative need, our results provided evidence for least effort. Nonetheless, we did not find evidence that the novelty of a new meaning predicts word-formation strategy.

The lack of evidence for the novelty hypothesis may have originated from two limitations. The first one is the size of our sense dataset, which only contains 109 English senses. In our analysis of relative novelty, we assumed that the senses we collected are representative of the full range of novelty, so that the preference for expressing high-novelty meanings with compounding may distinguish compound and reuse meanings. However, it may be the case that most meanings in our dataset have relatively low novelty. In this case, our hypothesis does not distinguish the two strategies, since both compounding and reuse may transparently express the emerging meaning. Another possibility is that high novelty disfavours compounding as well. Previous work suggests that the plausibility of a concept is crucial in compound interpretation (Costello & Keane, 2000). If high novelty implies implausibility, then expressing a high-novelty concept with compounds may harm communication, making compounding no more efficient than reuse.

Although we motivated communicative need and novelty (or the inverse of semantic similarity) as separate predictors of strategy choice, these two factors may jointly shape the label of an emerging meaning. For example, consider meanings with low need and high novelty, and observe that the word forms used to express them will also be infrequent and semantically opaque. Since infrequent forms can be more easily forgotten and replaced (Bybee, 2006) but transparent forms are more easily retained (Floyd & Goldberg, 2021; Brusnighan & Folk, 2012), infrequent and opaque existing forms may be more likely replaced by more transparent compound words. Another possibility is that when a high-need emerging meaning tends to co-occur with high-need and similar existing meanings, reusing similar words may actually harm informativeness and become an unfavourable strategy (Karjus, Blythe, Kirby, Wang, & Smith, 2021).

Our methodology builds on recent work in natural language processing that utilizes large historical corpora to examine changes in word meaning (e.g., Ryskina, Rabinovich, Berg-Kirkpatrick, Mortensen, & Tsvetkov, 2020; Hu et al., 2019). In this line of work, most closely related to ours is the study by Ryskina et al. (2020). Their work showed that emerging meanings can be differentiated from existing meanings by the density and rate of change in communicative need within their semantic neighbourhoods. Our work extends their analysis by examining the difference between meanings expressed via reuse and compound meanings.

Conclusion

We presented an initial study on how strategies are chosen to express emerging meanings, a topic that has not been investigated rigorously in previous work on word formation. We connected strategy choice with existing theories of communicative and cognitive efficiency, and found that a pressure for least effort predicts the word formation strategy used to express new meanings. Future work may extend this study by refining the theoretical framework, considering other word formation strategies such as morphological derivation, and testing it with historical data at a larger scale.

Acknowledgments

We would like to thank anonymous reviewers for constructive comments on an earlier version of the paper. AX is funded partly by the U of T–UoM IRTG program. This work was supported by NSERC Discovery Grant RGPIN-2018-05872 and ARC Future Fellowship FT19010020.

References

Bruni, E., Tran, N.-K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of artificial intelligence research*, *49*, 1–47.

Brusnighan, S. M., & Folk, J. R. (2012). Combining contextual and morphemic cues is beneficial during incidental vocabulary acquisition: Semantic transparency in novel compound word processing. *Reading Research Quarterly*, *47*(2), 172–190.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 711–733.

Cook, P., Lau, J. H., McCarthy, D., & Baldwin, T. (2014). Novel word-sense identification. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers* (pp. 1624–1635).

Costello, F. J., & Keane, M. T. (2000). Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science*, *24*(2), 299–349.

Davies, M. (2002). *The corpus of historical American English (COHA): 400 million words, 1810-2009*. Brigham Young University.

Dressler, W. U. (2005). Word-formation in natural morphology. In *Handbook of word-formation* (pp. 267–284). Springer.

Ferrer-i Cancho, R., Bentz, C., & Seguin, C. (2022). Optimal coding and the origins of Zipfian laws. *Journal of Quantitative Linguistics*, *29*(2), 165–194.

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., & Ruppin, E. (2001). Placing search in context: The concept revisited. In *Proceedings of the 10th international conference on world wide web* (pp. 406–414).

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.

Gagné, C. L., Spalding, T. L., & Schmidtke, D. (2019). Ladec: the large database of English compounds. *Behavior research methods*, *51*(5), 2152–2179.

Geeraerts, D. (1997). *Diachronic prototype semantics: A contribution to historical lexicology*. Oxford University Press.

Grewal, K., & Xu, Y. (2021). Chaining algorithms and historical adjective extension. *Computational approaches to semantic change*, *6*, 189.

Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2022). From partners to populations: A hierarchical Bayesian account of coordination and convention. *Psychological Review*.

Hill, F., Reichart, R., & Korhonen, A. (2015). Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, *41*(4), 665–695.

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

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, *165*, 45–52.

Karjus, A., Blythe, R. A., Kirby, S., Wang, T., & Smith, K. (2021). Conceptual similarity and communicative need shape colexification: An experimental study. *Cognitive Science*, *45*(9), e13035.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(1), 113–114.

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

Marchand, H. (1960). *The categories and types of present-day english word-formation*. Wiesbaden, Germany: Otto Harrassowitz.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, G. B., ... Norvig, P. (2011). Quantitative analysis of culture using millions of digitized books. *science*, *331*(6014), 176–182.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2018, May). Advances in pre-training distributed word representations. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018).* Miyazaki, Japan: European Language Resources Association (ELRA). Retrieved from `https://aclanthology.org/L18-1008`

Mollica, F., Bacon, G., Zaslavsky, N., Xu, Y., Regier, T., & Kemp, C. (2021). The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, *118*(49).

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

Pimentel, T., Nikkarinen, I., Mahowald, K., Cotterell, R., & Blasi, D. (2021, June). How (non-)optimal is the lexicon? In *Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 4426–4438). Online: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2021.naacl-main.350` doi: 10.18653/v1/2021.naacl-main.350

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.

Reimers, N., & Gurevych, I. (2019, 11). Sentence-bert: Sentence embeddings using Siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing.* Association for Computational Linguistics. Retrieved from http://arxiv.org/abs/1908.10084

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

Ryskina, M., Rabinovich, E., Berg-Kirkpatrick, T., Mortensen, D. R., & Tsvetkov, Y. (2020). Where new words are born: Distributional semantic analysis of neologisms and their semantic neighborhoods. *Proceedings of the Society for Computation in Linguistics*, *3*(1), 43–52.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in science conference* (Vol. 57, p. 61).

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

Štekauer, P. (2005). Onomasiological approach to word-formation. In *Handbook of word-formation* (pp. 207–232). Springer.

Xu, A., Kemp, C., Frermann, L., & Xu, Y. (2022). Word formation supports efficient communication: The case of compounds. In *Proceedings of the 44th annual meeting of the cognitive science society.*

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.

Yu, L., & Xu, Y. (2021). Predicting emergent linguistic compositions through time: Syntactic frame extension via multimodal chaining. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 920–931).

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