

# Chaining algorithms and historical adjective extension

Karan Grewal<sup>1</sup> and Yang Xu<sup>1,2</sup>

<sup>1</sup>Department of Computer Science, University of Toronto

<sup>2</sup>Cognitive Science Program, University of Toronto

This is the preprint for the following chapter (published version might differ):

Karan Grewal and Yang Xu. Chaining algorithms and historical adjective extension. To appear in: *Computational approaches to semantic change (Language Variation)*. Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu, & Simon Hengchen. Berlin: Language Science Press.

## Abstract

Natural language relies on a finite lexicon to express a potentially infinite set of ideas. This tension often results in the innovative reuse of existing words to describe emerging ideas. In this chapter, we take a computational perspective to examine how English adjectives extend their range over time to modify nouns and form previously unattested adjective-noun pairs. We hypothesize that how novel adjective-noun pairings emerge is non-arbitrary and follows a process of chaining, whereby novel noun referents for an adjective link to existing nouns modified by the same adjective that are close in semantic space. We test this proposal by exploring a set of probabilistic models that predict adjective-noun pairs from a historical text corpus (Google Books) that spans the past 150 years. Our findings across three diverse sets of adjectives support a chaining mechanism sensitive to local semantic neighbourhood – formulated as an exemplar model of categorization similar to the Generalized Context Model. These findings mirror existing work on chaining in the historical growth of grammatical categories. We discuss the limitations and implications of our approach toward a general theory of word meaning extension in natural language.

## 1 Introduction

Natural language relies on a finite lexicon to express a potentially infinite set of ideas. One result of this tension is the innovative reuse of existing words (Ramiro et al., 2018). Here we explore how English adjectives extend their range over time to modify novel nouns and ask whether there are principled mechanisms in the historical process of adjective extension.<sup>1</sup>

The topic of adjective-noun composition has been discussed in the computational literature. Existing studies have explored what adjective-noun pairings are considered plausible (Lapata et al., 1999), and how adjectives can be combined with nouns sensibly either via probabilistic models (Lapata, 2001) or through ontological constraints (Schmidt et al., 2006). Recent work has also suggested that adjective-noun composition can be modelled using vector-space models such as Word2Vec (Mikolov et al., 2013). In these studies, adjectives are considered to be linear operators that act on nouns in a vector space that impose linear transformations (Baroni and Zamparelli, 2010; Boleda et al., 2013; Vecchi et al., 2013, 2017) or conform to additive compositional models (Zanzotto et al., 2010). Despite this extensive line of work, sparse computational research has considered the dimension of time in the investigation of adjective-noun composition.

---

<sup>1</sup>See Grewal and Xu (2020) for a shorter conference version of this work.

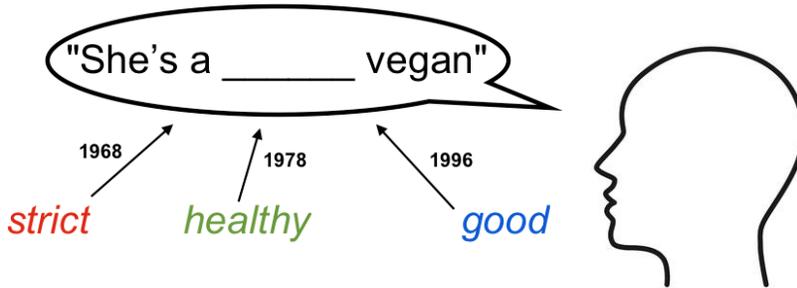


Figure 1: Example adjectives that emerged to describe *vegan* over the past half century.

Independent research in historical linguistics has explored adjective extension from the perspective of semantic change. In particular, Williams (1976) studied meaning change in synaesthetic adjectives and found that sensory terms such as those pertaining to sound, touch, and smell exhibit regular semantic shift such that words from the same sensory domain tend to undergo parallel change in meaning. For instance, Williams (1976) showed how adjectives that originally described the sense of touch have since extended to describe color (e.g., *warm cup* → *warm color*), and adjectives that originally described color have later extended to describe ideas associated with sound (e.g., *clear blue* → *clear voice*). This line of inquiry takes an empirical approach to characterize meaning change in adjectives from a focused semantic domain, but to our knowledge the more general problem of how adjectives extend their range to describe novel noun referents has not been treated formally or explored at scale.

We investigate whether adjective extension might follow non-random processes that make novel adjective-noun pairings yet to emerge in a linguistic community predictable. Our view is that novel adjective-noun pairings provide an incremental way of extending the referential range of adjectives, and word meaning extension or semantic change might result from this process (e.g., consider meaning extension in the adjective *cold* reflected in a chain of different noun context: *cold food* → *cold person* → *cold war*). It is conceivable that pairing with novel nouns does not necessarily entail semantic change in an adjective (e.g., *cold Gatorade* does not entail semantic change in *cold* which had the meaning ‘low-temperature’, even though *Gatorade* might appear as a novel item to pair with *cold* at some point in history), and our main focus here is to characterize the general mechanisms of an adjective’s extension over time – with or without semantic change.

Figure 1 illustrates that historical adjective-noun pairings can often be subject to non-linguistic or external influences which make them non-trivial to predict. For instance, the emergence of *vegan* is largely a cultural product, and different adjectives have been extended to modify this noun over time presumably as a result of cultural development. Our premise is that despite the historical adjective-noun pairings may be subject to socio-cultural influences, language users must somehow choose adjectives sensibly to describe nouns such that the novel pairings can be related to the original meaning of the adjectives. For this reason, we expect the historical processes of adjective extension to follow non-arbitrary paths.

We formulate adjective extension as a temporal prediction problem: Given adjective-noun pairings at historical time  $t$ , can we predict novel adjective-noun pairings into the future at  $t + \Delta$ ? We ground our work in cognitive linguistic theories of chaining, which have been proposed and recently demonstrated as important cognitive mechanisms for historical word meaning extension (Lakoff, 1987; Malt et al., 1999; Bybee et al., 1994; Sloman et al., 2001; Xu et al., 2016; Ramiro et al., 2018; Habibi et al., 2020). A consistent finding from these studies is that chaining as an extensional mechanism depends on semantic neighbourhood density, highlighting the fact that historical word meaning extension tends to follow incremental as opposed to abrupt processes. In our study, we consider each adjective as a linguistic category and explore different mechanisms of chaining to predict how adjective categories grow to modify

nouns that they have not previously been paired with. We next describe the theory of chaining and related work on word meaning extension.

## 2 Theory of chaining and word meaning extension

The proposal of semantic chaining is rooted in cognitive linguistic work on categories, or more specifically, radial categories (Lakoff, 1987). By this view, chaining is a process of meaning extension whereby novel items link to existing items of a linguistic category due to proximity in semantic space. This process leads to chain-like semantic structures, and Lakoff (1987) has considered it a key mechanism for growing radial categories or semantic networks, i.e., how categories grow ‘spokes’ of meaning from a central core meaning. Lakoff (1987)’s original work discusses chaining in a number of exemplary domains such as the grammatical categories of classifiers in Japanese and Dyirbal (an Australian aboriginal language), and prepositions such as how the English spatial term *over* extends over a wide variety of spatial (e.g., *over the hill*) and metaphorical context (e.g., *over the moon*). Later work also discusses chaining in the grammar evolution of tense, modality, and aspect systems (Bybee et al., 1994), container naming (Malt et al., 1999), and metonymical semantic shift (Hilpert, 2007). These studies have broadened the view of chaining toward a generic mechanism for grammatical and semantic changes in language, although they do not provide a formal account for the processes of chaining or test this idea comprehensively against historical corpus data.

Extending the cognitive linguistic accounts of chaining, recent work has explored formal approaches to chaining in several aspects. Sloman et al. (2001) and Xu et al. (2016) have developed computational models of chaining and tested the extent to which these models account for the extension of container names such as *bottle* and *jar*. Their findings suggest that chaining depends on semantic neighbourhood density, and more specifically nearest-neighbour models of chaining tend to best account for the empirical data. Ramiro et al. (2018) extends this work to examine whether similar models of chaining might explain the historical emergence of senses (or word sense extension) in English words over the past millennium, e.g., how *face* might extend from ‘body part’ to senses including ‘front surface (of an object)’, ‘facial expression’, and ‘defy danger’. Their work confirms the earlier finding that chaining relies on semantic neighbourhood density, and senses tend to emerge by linking those that are close in semantic space.

More recent work has built on these computational studies to investigate the historical growth of grammatical categories, and particularly numeral classifiers commonly used in East Asian languages (Habibi et al., 2020). This work has examined a suite of probabilistic models of chaining and found chaining to be best captured by an exemplar model, also known as the Generalized Context Model in the psychological literature of categorization (Nosofsky, 1986). By this view, chaining in linguistic categories reflects an exemplar-based process of extension that mirrors those found in other aspects of language change including phonetics, morphology, word senses, and constructions (Skousen, 1989; Pierrehumbert, 2001; Keuleers, 2008; Bybee, 2013; Ramsey, 2017).

Here we examine chaining through the lens of the exemplar theory but in a new domain: the case of historical adjective extension in English. Analogous to how numeral classifiers (e.g., in Mandarin Chinese) extend toward novel nouns, English adjectives also extend to modify novel noun referents. If the exemplar view represents a general mechanistic account for the growth of linguistic categories, it should explain the historical extension of adjective categories.

Figure 2 illustrates the exemplar theory of chaining with two example adjectives and a dimension-reduced semantic space of their noun referents, data for which were taken from the Google Books corpus (Michel et al., 2011) during the 1880s. The two adjectives *wrong* and *troubled* are closely related in semantic space in the 1880s and share noun referents (labelled in purple) such as *war* and *humanity*. The emergent or query noun *slavery* has not appeared in close context with either adjective prior to the 1880s but is in semantic proximity of their noun referents. The exemplar view of chaining postulates that the linguistic category having a higher local semantic similarity (or neighbourhood density) to a novel referent is more likely to attract that item, and when this process repeats over time chain-like category structures may result in semantic space. Here, *wrong* has a higher neighbourhood density (with its noun

referents labelled in red) to *slavery* in comparison to *troubled* (with its noun referents labelled in blue), namely that the existing noun referents of *wrong* are closer in semantic space to the query noun than those of *troubled*. The exemplar view of chaining thus predicts that *wrong* is a more likely adjective candidate to be paired with *slavery*, which aligns with the empirical data. We seek to evaluate the extent to which the exemplar model of chaining accounts for historical adjective extension, and if it is better or worse than alternative accounts for the chaining process.

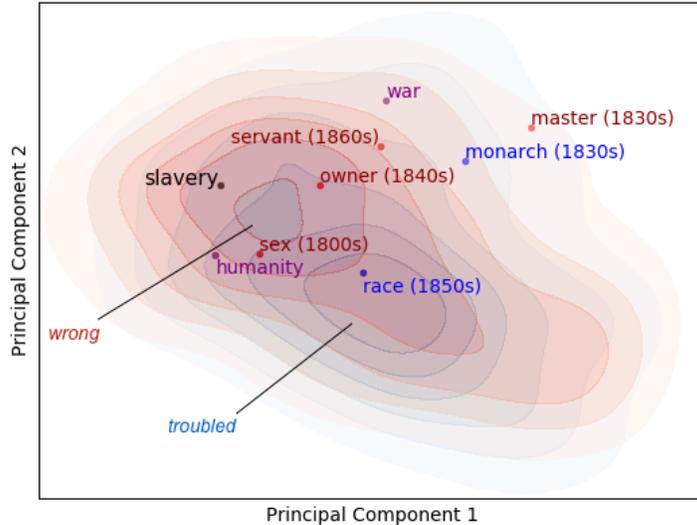


Figure 2: An illustration for the exemplar view of semantic chaining (Habibi et al., 2020) using two example adjectives *wrong* and *troubled*. The semantic space is constructed from the first 2 principal components in the Principal Components Analysis on the diachronic Word2Vec embeddings from the 1870s (Hamilton et al., 2016). Nouns labelled in purple (e.g., *humanity*, *war*) are shared context of the two adjectives. Nouns labelled in red (e.g., *master*, *servant*, *owner*, *sex*) and blue (e.g., *monarch*, *race*) are contexts that co-occurred more often with *wrong* and *troubled* respectively up to the 1880s. The contours represent probability distributions of nouns co-occurring with each of the two adjectives, constructed by kernel density estimation.

### 3 Computational formulation of theory

We formulate adjective extension as a temporal categorization problem and explore the process of chaining via a suite of models that predict adjective-noun pairings over time. The probabilistic formulation we describe here follows existing work on chaining and the extension of numeral classifiers (Habibi et al., 2020).

#### 3.1 Probabilistic formulation

Given an emergent query noun  $n^*$  at a future time  $t + \Delta$  and a finite set of adjectives  $\mathcal{A}$ , we seek to predict which adjective(s)  $a \in \mathcal{A}$  would be most appropriate for describing  $n^*$  at time

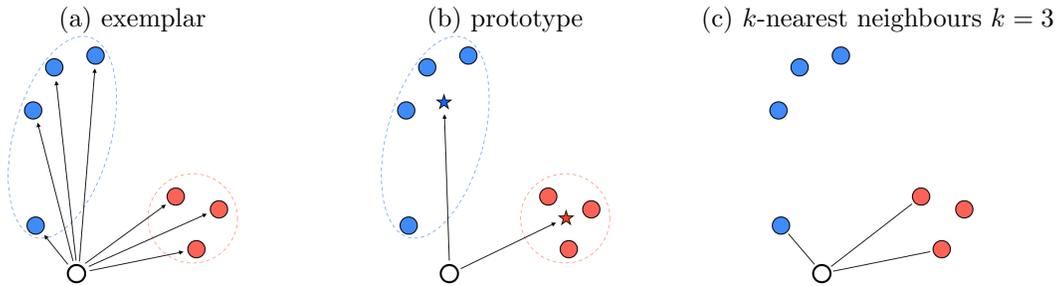


Figure 3: An illustration of representative chaining models for the likelihood function. The empty circle represents the stimulus or the query noun  $n^*$ . Red circles represent nouns that are attested to have paired with one particular adjective, and blue circles represent nouns that are attested to have paired with an alternative adjective (in reality, a noun can pair up with multiple adjectives). The dotted lines indicate the noun referent space for a given adjective. The stars represent the prototypes under the prototype model. The lines indicate the influence of existing (exemplar) nouns to the query noun as specified in each model of chaining.

$t + \Delta$  based on the historically attested adjective-noun pairings at current time  $t$ .<sup>2</sup> We cast this problem as probabilistic inference over the space of adjectives for a query noun  $n^*$ :

$$p(a|n^*)^{(t+\Delta)} \propto p(n^*|a)^{(t)} p(a)^{(t)}. \quad (1)$$

The posterior term  $p(a|n^*)^{(t+\Delta)}$  relies on two sources of information to predict the choice of adjective(s) for  $n^*$ : (1) a likelihood function  $p(n^*|a)^{(t)}$  that specifies the semantic proximity of  $n^*$  to an adjective  $a$  given knowledge of its existing noun referents at time  $t$ , and (2) a prior distribution  $p(a)^{(t)}$  that captures the *a priori* belief or probability of choosing an adjective  $a$  from the current lexicon without considering its semantic relation to  $n^*$ . In both our formulations of the likelihood and the prior, we focus on type-based representations of adjective-noun co-occurrence frequencies and adjective frequencies. Token-based representations have been explored and shown to be inferior in accounting for the historical growth of classifier categories in related recent work (Habibi et al., 2020).

## 3.2 Likelihood function

We describe a suite of models to explore a space of possible candidates for the likelihood function. Each of these models postulates a different mechanism of chaining that links existing noun referents of an adjective to a novel noun that appears at a future time. We use  $\{n\}_a^{(t)}$  to denote the semantic embeddings for the set of nouns that co-occur with adjective  $a$  at current time  $t$ , i.e., the semantic representation for the collective set of noun referents for adjective category  $a$ . Figure 3 provides an illustration for the representative chaining models that we describe in the following subsections.

### 3.2.1 Exemplar model

The first likelihood function we consider is based on the exemplar theory which is discussed in the psychological literature of categorization (Nosofsky, 1986). Here each noun  $n \in \{n\}_a^{(t)}$  is treated as an exemplar for an adjective  $a$ .

The exemplar view of chaining postulates that a query noun should be linked to an adjective category where the noun exemplars are most proximal in semantic space. As such, a novel noun is pulled or attracted to the adjective category that has the highest local semantic density around that noun. The likelihood term between  $n^*$  and adjective  $a$  is thus proportional to the weighted sum of similarities between  $n^*$  and the noun exemplars of  $a$ :

<sup>2</sup>In our formulation of the prediction problem, we consider an adjective-noun pair to be novel if 1) the noun itself is novel or 2) the pairing has not been attested in history.

$$p(n^*|a)^{(t)} \propto \frac{1}{h \left| \{n\}_a^{(t)} \right|} \sum_{n \in \{n\}_a^{(t)}} \text{sim}(n^*, n) \quad (2)$$

The similarity function  $\text{sim}(\cdot, \cdot)$  measures how similar two nouns are and is defined as the exponentiated negative distance in semantic space which assigns differential weights to exemplars based on their relative distances to the query (higher similarities for more proximal exemplars):

$$\text{sim}(n^*, n) = e^{-\frac{d(n^*, n)^2}{h}} \quad (3)$$

$d(\cdot, \cdot)$  measures the Euclidean distance between nouns and  $h$  is a kernel parameter that we learn from data. The choice of the exemplar model and the similarity formulation is grounded in work on Generalized Context Model (Nosofsky, 1986), which has recently been shown to predict the historical extension of Chinese numeral classifiers (Habibi et al., 2020). Here we examine whether the same exemplar-based processes of chaining might explain historical adjective extension. This model is also equivalent to performing kernel density estimation in semantic space defined by the likelihood function, and thus we use a kernel parameter  $h$  in the similarity function and normalize the term by dividing the resulting sum by  $h$ .

### 3.2.2 Prototype model

Motivated by earlier psychological work on prototype theory (Rosch, 1975) and related recent work on few-shot learning (Snell et al., 2017), we consider an alternative view of chaining based on category prototypes. Each adjective  $a$  is represented by a prototype at time  $t$  that captures the ‘gist’ of noun referents for that category. We operationalize the prototype as the expectation of all exemplars within a category:

$$\bar{p}_a = \mathbb{E} \left[ n \in \{n\}_a^{(t)} \right] = \frac{1}{\left| \{n\}_a^{(t)} \right|} \sum_{n \in \{n\}_a^{(t)}} n \quad (4)$$

The likelihood function postulates a chaining mechanism that links a query noun to the adjective that has the closest prototype in semantic space:

$$p(n^*|a)^{(t)} \propto \text{sim}(n^*, \bar{p}_a) = e^{\left( -\frac{d(n^*, \bar{p}_a)^2}{h'} \right)} \quad (5)$$

Similar to the exemplar model, we use a kernel parameter  $h'$  that controls how quickly similarity scales with respect to the semantic distance between the query noun and the prototype. This model can behave differently from the exemplar model of chaining: even if a query noun is closer to the prototype of one adjective over an alternative adjective, a small set of exemplars closest to that noun can pull the query item to the alternative category (see Habibi et al., 2020 for a simulation that compares the properties of the exemplar and prototype models of chaining).

We also consider a variant of the prototype model in which the prototype representation for each adjective category remains static over time. That is,  $\bar{p}_a = \bar{p}_a^{(t_0)}$  for all  $t > t_0$  where  $t_0$  is the initial time of investigation. We refer to this variant as the progenitor model.

### 3.2.3 $k$ -nearest neighbours model

In addition to the exemplar and prototype models, we consider a family of models based on  $k$ -nearest neighbours ( $k$ -NN). In a Bayesian framework, the  $k$ -NN likelihood of  $n^*$  pairing up with adjective  $a$  is proportional to whether its  $k$  closest neighbours  $n_1, \dots, n_k$  previously paired up with  $a$ , and inversely proportional to the size of category  $a$ :

$$p(n^*|a)^{(t)} \propto \frac{1}{\left| \{n^*\}_a^{(t)} \right|} \sum_{j=1}^k I(n_j \in \{n\}_a^{(t)}) \quad (6)$$

Here the sum is over the  $k$  nouns closest to  $n^*$  in semantic space. When this likelihood is combined with the prior, the  $k$ -NN posterior probability amounts to  $n^*$ 's  $k$  closest neighbours voting for each of the adjectives that they previously paired up with.

This formulation of  $k$ -NN can be viewed as a ‘hard version’ of the exemplar model where  $k$  is a discrete analog of the kernel parameter  $h$ . We report  $k = 1$  and  $k = 10$  in our experiments.

### 3.3 Prior distribution

We formulate a type-based prior  $p(a)^{(t)}$  which specifies how likely adjective  $a$  is to be paired with any noun based on the set size of its noun referents at time  $t$ . This prior formulation predicts that  $a$ 's probability of appearing in a novel adjective-noun pairing is directly proportional to the number of unique nouns it has previously paired up with:

$$p(a)^{(t)} = \frac{|\{n\}_a^{(t)}|}{\sum_{a' \in \mathcal{A}} |\{n\}_{a'}^{(t)}|} \quad (7)$$

The rationale behind this choice of prior is as follows: if semantic chaining underlies the emergence of novel adjective-noun pairs, then adjectives that have paired with more nouns would have a higher *a priori* probability of attracting a query noun  $n^*$  via linking it to semantically similar nouns which are more likely to have previously co-occurred with  $a$  (Luo and Xu, 2018). This rich-get-richer process is also supported by work on how semantic networks grow through preferential attachment (Steyvers and Tenenbaum, 2005).

This category-size-based prior serves as our baseline model when making adjective predictions for  $n^*$  at time  $t + \Delta$ , where  $p(a|n^*)^{(t+\Delta)} = p(a)^{(t)}$ . We focus on the type-based representation as opposed to token frequencies because work from Habibi et al. (2020) has shown that a type-based prior worked better than a token-based prior in predicting the extension of grammatical categories.

### 3.4 Semantic space

To construct a semantic space for the nouns, we use word embeddings, particularly Word2Vec, commonly used for distributed semantic representation in natural language processing (Mikolov et al., 2013). We choose this construction of semantic space partly because it has been demonstrated to be effective in predicting grammatical category extension (Habibi et al., 2020). However, adjective usage is likely to entail a semantic representation richer than purely linguistic information, and future work should explore alternative methods for constructing semantic space such as those based on perceptual features and lexical taxonomic structures.

Since the word co-occurrence distributions are constantly changing over time, our semantic representations (of nouns) also need to be updated accordingly. For this reason, we use diachronic (or historical) Word2Vec embeddings (Hamilton et al., 2016) where at each time  $t$ , the embedding for a noun is based on its co-occurrence profiles at time  $t$ , relatively independent to future co-occurrences. In this respect, the predictions made by our models are in some sense ‘zero-shot’, or deprived of semantic information into the future.

## 4 Data

We extracted a large database of historical adjective-noun pairings over the past 150 years (1850–2000). We collected these data from the Google Books corpus (Michel et al., 2011) which contains sentence fragments from historical books over the past five centuries. Within Google Books, the English All (ENGALL) corpus accounts for  $8.5 \times 10^{11}$  tokens and roughly 4% of all books ever published. Given the size of the ENGALL corpus, is likely to reflect how the English language has changed over the past centuries, making our adjective-noun co-occurrence dataset suitable for evaluating hypotheses about chaining.

We collected adjective-noun co-occurrence counts from the ENGALL corpus. First, we extracted all bigrams from the ENGALL corpus in which the first token is an adjective and the second is a noun (by POS tags specified in the data) along with the corresponding timestamp.

Since the corpus is likely to contain noise, we standardized the set of nouns and adjectives by only considering those present in WordNet (Miller, 1995), which yields approximately 67k nouns and 14k adjectives.

We collapsed raw co-occurrence counts into decadal bins by choosing  $\Delta = 10$  years. This yielded our adjective-noun pairings dataset which consists of entries of the form  $(a, n, \text{count}, t)$ . In each decade  $t$ , we used a Word2Vec language model pre-trained on historical text (i.e., digitized books from Google Books) for the semantic representation. For our analyses, we worked with a subset of the collected data (discussed in the next section), due to both considerations of sampling diversity and computational feasibility. To construct semantic representations across decades, we used diachronic Word2Vec embeddings which were trained using the EN-GALL corpus. (Hamilton et al., 2016) also chose to construct diachronic Word2Vec embeddings decade-by-decade for similar reasons.

We now describe three adjective sets  $\mathcal{A}$ . The purpose of evaluating our models on three different adjective sets is to obtain representative samples of the adjectives, and to ensure our hypotheses are robust to the choice of adjectives.

1) *Frequent adjectives*. We use multiple ways to construct  $\mathcal{A}$  such that it covers a broad scope and we show our results are reproducible and agnostic to choice of adjectives. To construct a set of 200 adjectives that cover a broad range of descriptions, we first collected word vectors of all adjectives in the Google Books corpus using a pre-trained Word2Vec model. Next, we clustered the adjectives into 20 clusters and picked 10 adjectives from each to construct our set  $\mathcal{A}$  of 200 adjectives. We applied this clustering procedure to obtain a feasibly large yet diverse set of adjectives for the analyses, and we used the  $k$ -means algorithm for clustering. Adjectives were sampled from each cluster based on their usage frequencies, and only considered against other adjectives within the same cluster during sampling. We refer to this set as FRQ-200, with examples shown in Table 1.

2) *Random adjectives*. To ensure that the sampling scheme for choosing  $\mathcal{A}$  is not biased towards token frequencies, we also constructed another set of 200 adjectives by repeating the clustering step described above, but we replaced frequency sampling with uniform sampling. We refer to this dataset as RAND-200. As Table 1 shows, adjectives drawn from the same cluster are semantically similar between FRQ-200 and RAND-200, but less common in the latter set.

3) *Synaesthetic adjectives*. We also consider the third set of *synaesthetic adjectives* (SYN-65) defined by Williams (1976), as a more focused domain that is known to undergo semantic change. This set includes 65 adjectives that exhibit regular semantic shift historically. We will refer to this set as SYN-65.<sup>3</sup>

Data and code from our analyses are available at:

<https://github.com/karangrewal/adjective-extension>

Table 1: A comparison of some adjectives in FRQ-200 and RAND-200 grouped according to the cluster they were drawn from. Notice that the clusters (per column) align semantically, however the adjectives in FRQ-200 are more frequently represented in the English lexicon than those in RAND-200.

FRQ-200	RAND-200	FRQ-200	RAND-200
<i>Asian</i>	<i>Hungarian</i>	<i>polite</i>	<i>chatty</i>
<i>Christian</i>	<i>Thai</i>	<i>intelligent</i>	<i>unorthodox</i>
<i>American</i>	<i>Cornish</i>	<i>passionate</i>	<i>amiable</i>
<i>European</i>	<i>Catalan</i>	<i>energetic</i>	<i>communicative</i>

<sup>3</sup>There are in fact 64 unique adjectives in this set and WordNet captures 61 of these adjectives. See Williams (1976) for a comprehensive list.

## 5 Results

We present results in two steps. First, we examine the set of chaining algorithms described on novel adjective-noun pairings that appeared during 1850-2000, and we evaluate whether the exemplar model better predicts these data than the alternative models. Second, we perform a more focused analysis to examine whether the chaining algorithms predict extensional patterns in adjectives that show most and least semantic change over the past 150 years.

### 5.1 Evaluation of chaining algorithms

We evaluated the set of chaining algorithms on their ability to predict which adjectives  $a \in \mathcal{A}$  would pair up with a given noun  $n^*$  in decade  $t + \Delta$  given information about  $n^*$  up to and including decade  $t > t_0$ , where  $t_0$  is the base decade. This information includes co-occurrences between all nouns  $n$  and adjectives  $a \in \mathcal{A}$  at or before decade  $t$ , as well as time-dependent word embeddings at each decade taken from (Hamilton et al., 2016). We chose  $t_0$  as the 1840s and built a base lexicon from adjective-noun co-occurrences between  $t_0$  and the 2000s. The 1860s was the first decade for which we report model prediction, and we used the 1850s as our ‘training decade’ to estimate the kernel parameters for the exemplar and prototype models.

We define pairings  $(a, n^*)$  to be novel in decade  $t + \Delta$  if and only if (i)  $a$  co-occurred with  $n^*$  in decade  $t + \Delta$  beyond a certain threshold (which we set to 2), and (ii)  $a$  never appeared with  $n^*$  beyond that threshold in any decade  $t' < t$ . Using these criteria allowed us to eliminate noise from co-occurrence statistics. Given a noun  $n^*$ , each model’s output was a categorical distribution  $p(a|n^*)^{(t+\Delta)}$  over all adjectives  $a \in \mathcal{A}$ . The model was then scored on its precision accuracy on the set of adjectives that first co-occurred with  $n^*$  in decade  $t + \Delta$ . That is, if  $n^*$  co-occurred with  $m$  new adjectives in  $\mathcal{A}$  in decade  $t + \Delta$ , we took the top  $m$  adjectives with the highest posterior probabilities that had not previously co-occurred with  $n^*$  as the set of retrieved positives. This evaluation metric calculates the percentage of correct predictions a model makes, and it is identical to the metric used in previous work for the prediction of historical extensions of classifier categories (Habibi et al., 2020). We report the total precision for all models and use this metric as an objective function to learn the kernel parameters from the initial training decade. We consider two types of predictive tasks when making predictions for noun  $n^*$  in decade  $t$ : taking as ground truth adjectives that co-occur with  $n^*$  (1) specifically in the immediate future decade  $t + \Delta$ , and (2) all future decades  $t' > t$  up to the terminal decade 1990s.

We summarize results from our experiments for the three differently sampled adjective sets  $\mathcal{A}$ . As Figure 4 shows, the exemplar model has the highest predictive performance, followed closely by the 10-NN and prototype models. The exemplar, prototype, and 10-NN models perform substantially better than the baseline. These results provide evidence that chaining may rely on mechanisms sensitive to semantic neighbourhood density, best captured by the exemplar model. We also observed that the 10-NN model did not perform better than the exemplar model as the kernel parameter is a continuous analog of  $k$  and is optimized for precision, but increasing  $k$  in  $k$ -NN beyond 10 did help to improve model prediction suggesting that local neighbourhood density matters in predicting adjective extension. The progenitor model, a variant of the prototype model with static prototypes determined in decade  $t_0$ , is considerably worse than the prototype model with a moving prototype. This relationship between the prototype and progenitor models that we observe indicates that if the prototype model is the closest underpinning of adjective extension, then  $\{n\}_a^{(t)}$  largely influences which nouns adjective  $a$  will extend to and that each adjective category ‘center’ updates once novel adjective-noun pairings are formed. We also observed that the baseline or prior model performed worse than the exemplar and prototype models, suggesting that semantic relations matter in adjective-noun pairing, above and beyond the size-based adjective priors.

Further results with year-over-year accuracy breakdowns are shown in Figure 5. The predictive accuracy falls in later decades since there are fewer novel adjective-noun pairings to predict. Our results hold generally across the three adjective sets, and they suggest that semantic neighbourhood density is an important factor contributing towards adjective extension as the exemplar and 10-NN models achieve overall better predictive accuracy over the other alternative models.

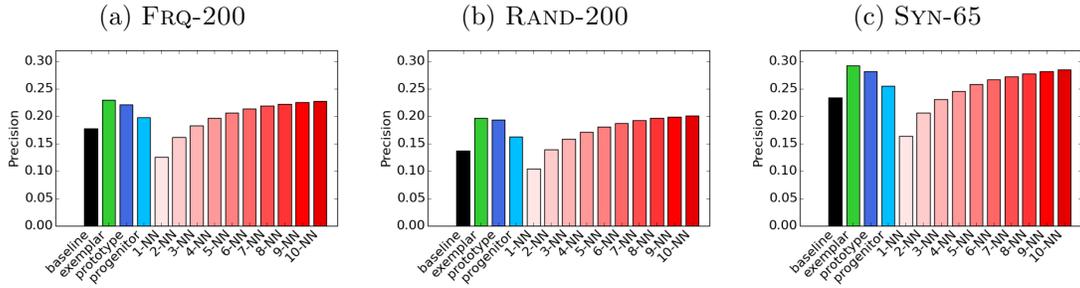


Figure 4: Aggregate precision accuracy for all models (including  $k$ -NN from  $k = 1$  to  $k = 10$ ) across all time periods on each of our three adjectives sets.

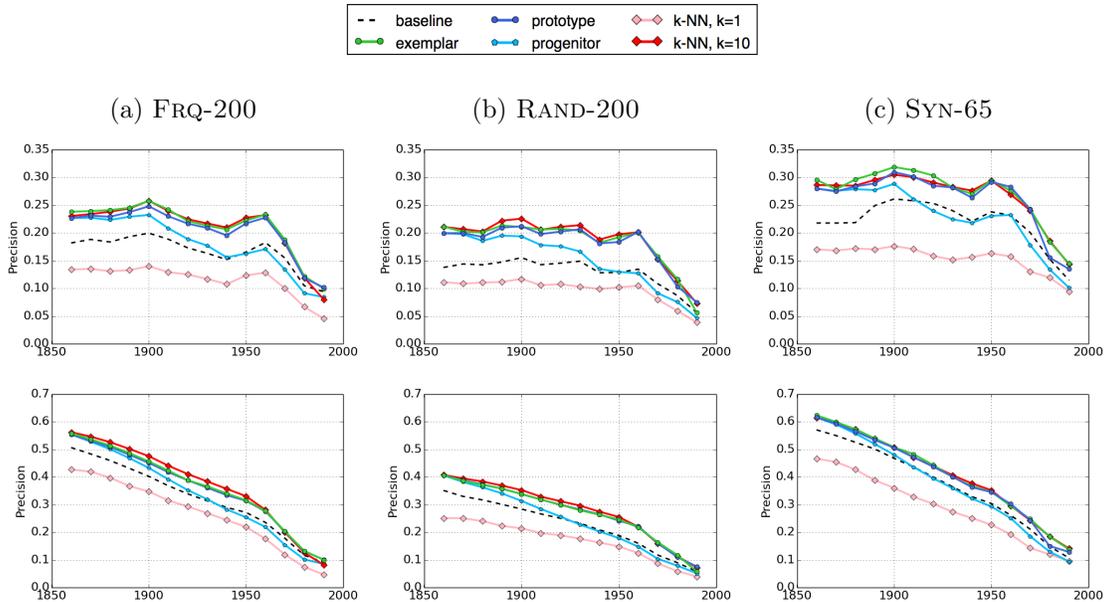


Figure 5: Model predictive accuracy on the FRQ-200, RAND-200, and SYN-65 adjective sets. *Top row*: Predictive accuracy when only novel adjective-noun pairs in the following decade are considered. *Bottom row*: Predictive accuracy when all future adjective extensions are considered.

Table 2 provides some examples of model prediction and highlights the limitations of the approach. It is worth noting that while the exemplar model performed well in comparison to the other models, all models failed to predict parts of the empirical data. This issue might be partly due to the fact that our semantic representation of nouns is inadequate to capture the kinds of rich knowledge that determines adjective modification of nouns, and partly due to the historical events that add randomness to the process, e.g., how alcohol prohibition in the 1920s made *illegal* an appropriate adjective modifier for *alcohol*, and how *American* and *Vietnam* became associated in context presumably due to the Vietnam War around the 1960s.

## 5.2 Chaining in semantically changing and stable adjectives

We next examine the extent to which the chaining algorithms predict extensional patterns in both semantically changing and stable adjectives in history. Because the chaining view presumes meaning change to take incremental (as opposed to abrupt) steps, it is plausible that it is less effective in predicting adjective extension in those adjectives that show substantial

Table 2: Examples of model prediction on the FRQ-200 adjective set. Adjectives with an asterisk (\*) indicate true positives retrieved by models. We present predictions for nouns *cigarette*, *alcohol*, and *Vietnam* as the adjectives they first pair with in the 1880s, 1920s, and 1960s respectively reflect sentiment (e.g., *social cigarette*) or historic events (e.g., *illegal alcohol* due to prohibition, *American Vietnam* due to the Vietnam war).

noun & decade	<i>cigarette</i> , 1880s
new adjectives	<i>better</i> , <i>modern</i> , <i>several</i> , <i>excessive</i> , <i>American</i> , <i>social</i>
baseline prediction	<i>original</i> , <i>particular</i> , <i>English</i> , <i>natural</i> , <i>perfect</i> , <i>modern*</i> (1/6)
exemplar prediction	<i>black</i> , <i>red</i> , <i>English</i> , <i>poor</i> , <i>original</i> , <i>particular</i> (0/6)
prototype prediction	<i>red</i> , <i>black</i> , <i>dry</i> , <i>warm</i> , <i>cold</i> , <i>English</i> (0/6)
10-NN prediction	<i>original</i> , <i>warm</i> , <i>particular</i> , <i>red</i> , <i>English</i> , <i>dry</i> (0/6)
noun & decade	<i>alcohol</i> , 1920s
new adjectives	<i>female</i> , <i>analogous</i> , <i>red</i> , <i>bitter</i> , <i>marked</i> , <i>illegal</i>
baseline prediction	<i>perfect</i> , <i>extraordinary</i> , <i>moral</i> , <i>physical</i> , <i>western</i> , <i>christian</i> (0/6)
exemplar prediction	<i>red*</i> , <i>moral</i> , <i>artificial</i> , <i>dense</i> , <i>perfect</i> , <i>marked*</i> (2/6)
prototype prediction	<i>artificial</i> , <i>perfect</i> , <i>marked*</i> , <i>red*</i> , <i>physical</i> , <i>moral</i> (2/6)
10-NN prediction	<i>red*</i> , <i>moral</i> , <i>dense</i> , <i>perfect</i> , <i>analogous*</i> , <i>artificial</i> (2/6)
noun & decade	<i>Vietnam</i> , 1960s
new adjectives	<i>western</i> , <i>tropical</i> , <i>eastern</i> , <i>colonial</i> , <i>particular</i> , <i>more</i> , <i>top</i> , <i>poor</i> , <i>American</i>
baseline prediction	<i>same</i> , <i>more*</i> , <i>great</i> , <i>particular*</i> , <i>American*</i> , <i>different</i> , <i>natural</i> , <i>human</i> , <i>English</i> (3/9)
exemplar prediction	<i>western*</i> , <i>eastern*</i> , <i>more*</i> , <i>particular*</i> , <i>great</i> , <i>colonial*</i> , <i>inner</i> , <i>same</i> , <i>poor*</i> (6/9)
prototype prediction	<i>great</i> , <i>same</i> , <i>western*</i> , <i>more*</i> , <i>American*</i> , <i>eastern*</i> , <i>particular*</i> , <i>European</i> , <i>French</i> (5/9)
10-NN prediction	<i>western*</i> , <i>eastern*</i> , <i>more*</i> , <i>tropical*</i> , <i>colonial*</i> , <i>great</i> , <i>better</i> , <i>inner</i> , <i>particular*</i> (6/9)

change in meaning over time. However, if chaining reflects a generic mechanism of meaning change, we should expect the models described to predict both semantically changing and stable adjectives.

To investigate this issue, we performed a group comparison where we split each adjective set into two subsets: a semantically changing group that showed the highest degrees of semantic change, and a semantically stable group that showed the least degrees of semantic change. We defined the degree of semantic change of an adjective based on its semantic neighbourhood profiles during the flanking decades: 1850s and 1990s. We followed the same procedure as Xu et al. (2015), where we calculated the degree of overlap in 100 semantic neighbours in adjectives (using diachronic word embeddings from Hamilton et al., 2016) between the flanking decades and took the inverse of that quantity as degree of change: a fully stable adjective would have 100% overlap in its neighbourhood, whereas a highly changing adjective would have low % overlap in its neighbourhood. We then applied the same models of chaining to these two subgroups in each of the three adjective sets.<sup>4</sup>

We analyzed the 50 top and least changing adjectives from the FRQ-200 and RAND-200 sets, and only 20 top and least changing adjectives from the SYN-65 set because it contained 61 adjectives in total. The results appear in Figure 6. We observed that the proposed algorithms of chaining, particularly the exemplar, prototype, and 10-NN models, perform substantially better than the frequency baseline. This observation holds for both the semantically changing and stable adjective subgroups, suggesting that chaining mechanisms apply equally to these

<sup>4</sup>For the prototype model, we only present results based on the (moving) prototype model because it was shown to be a superior model than the progenitor model that assumes the prototype to be time-invariant, both in Section 5.1 and Habibi et al. (2020).

adjective sets. In both the FRQ-200 and RAND-200 sets, the exemplar model consistently outperforms the alternative models in predictive accuracy over time, but its performance is not the strongest in the SYN-65 although this particular set has the smallest subset size (20 adjectives). These results suggest that chaining is a generic mechanism in historical adjective extension.

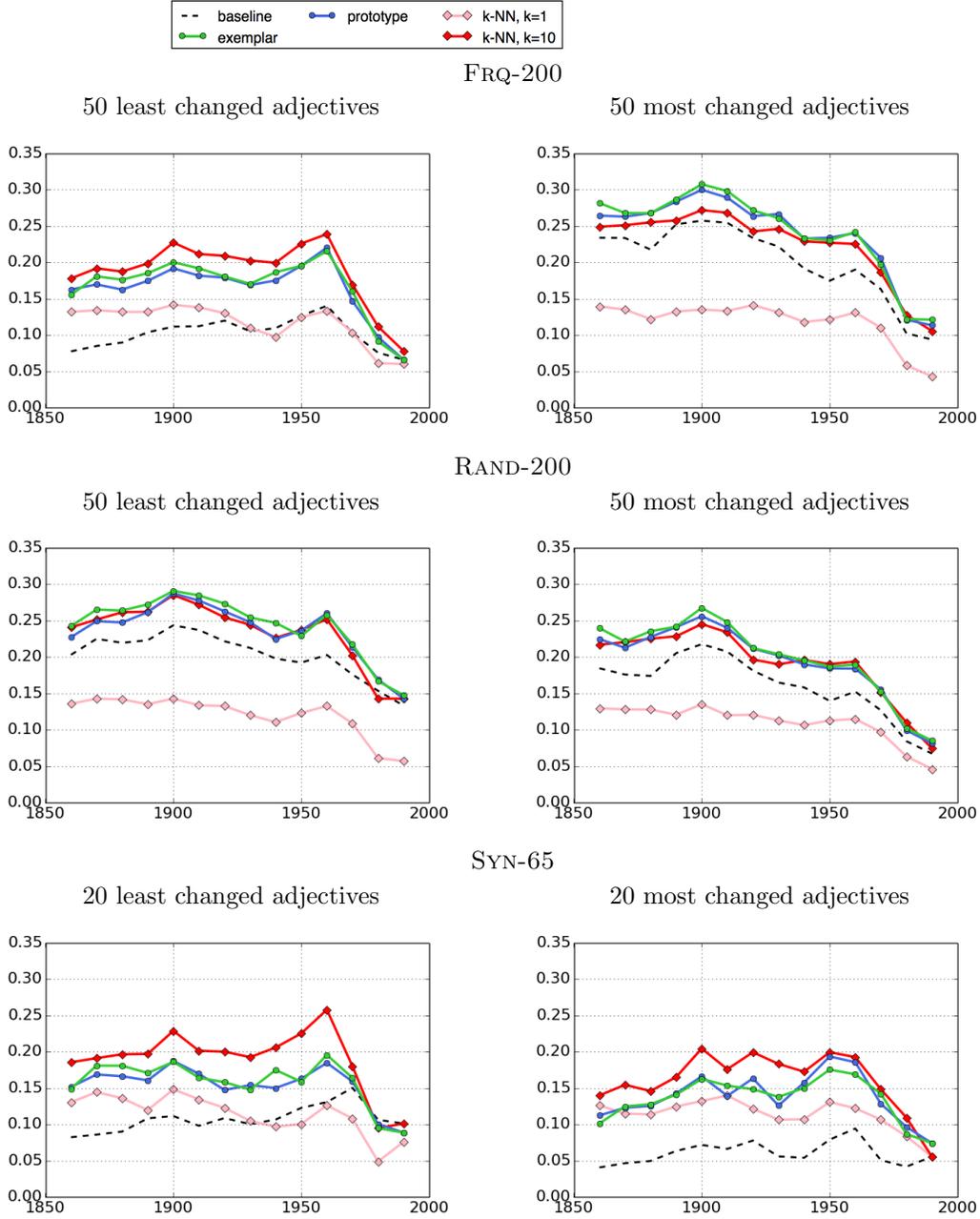


Figure 6: Model predictive accuracy on the most semantically changing and stable adjectives from FRQ-200, RAND-200, and SYN-65 adjective sets.

## 6 Discussion

Our findings support the overall hypothesis that semantic neighbourhood density influences how novel adjective-noun pairings emerge, although the distinction between the exemplar model and the alternative models is small for drawing strong conclusions from this initial investigation. Nevertheless, all the models we examined perform considerably better than the baseline model. Our work mirrors existing studies on chaining in the extension of grammatical categories (Lakoff, 1987; Bybee et al., 1994; Habibi et al., 2020), and we discuss its limitations and implications toward a general theory of word meaning extension.

### 6.1 Limitations

Our formulation of chaining depends on semantic similarity. One drawback of this assumption is that although chaining mechanisms may retrieve nouns that are similar to a query noun, there is no independent mechanism of checking whether the adjective-noun pairing is plausible. That is, our implementation of chaining does not explicitly ‘perform a check’ as to whether a predicted adjective-noun pairing is sensible. As adjectives accumulate novel senses, the set of possible nouns they can pair with will also vary due to external factors orthogonal to the internal mechanism of chaining. Here we acknowledge this limitation and consider it an important future direction to explore the interaction of internal and external factors that co-shape word meaning extension and semantic change.

Throughout our analyses we have assumed that distributed semantic representations, or word embeddings, are sufficient to capture the meaning of nouns. In particular, we used Word2Vec to capture distributional meaning of words from linguistic context, but other variants of semantic representation are available and should be considered in future explorations. Importantly, perceptual (e.g., visual) features might be especially relevant for constructing the meaning of concrete nouns, and our current construction of the semantic space might not capture these features. There exists computational work that explores adjective meaning using a combination of visual and linguistic information. For instance, Lazaridou et al. (2015) applied cross-modal mappings between visual and linguistic representations to assign adjective labels to visual inputs, and Nagarajan and Grauman (2018) followed up by learning a linear mapping that predicts adjective descriptors based on visual input. However, one limiting factor of these cross-modal approaches is that they may not be relevant to predicting adjective pairings with abstract nouns where perceptual grounding is more difficult to establish. In these cases, both socio-cultural factors and cognitive devices such as metaphor may be relevant in predicting adjective extension, above and beyond the semantic representation and the simple chaining mechanisms that we have considered.

Our analyses have relied on written text (i.e., books) which might not be fully representative of natural language use that also involves colloquialism and conversations (represented more accurately in spoken text corpora). The interpretations we drew from our analyses are thus restricted to formal forms of language, although they are also useful reflections of conventional language use. Earlier work by Williams (1976) on synaesthetic adjectives has also used dictionaries as a source of investigation, and a potential research direction is to examine the properties of adjective meaning extension or change in both written and spoken text. Written language is likely to be a delayed reflection of spoken language, and as such we might expect changes in word meaning and usage in spoken language to precede those in written text. Colloquialism may also add nuances beyond this difference, whereby language use is notably more casual and flexible (partly due to the socio-cultural knowledge involved), e.g., emergent adjective usages in slang might be harder to predict in comparison to the case of formal written text.

### 6.2 Relations of chaining and semantic change

The proposal of semantic chaining as initially described by Lakoff (1987) has focused on the formation of complex linguistic categories, particularly the grammatical classes such as classifiers and prepositions. Although Lakoff did not discuss extensively the relations of chaining and historical semantic change, the anecdotal cases that he described have assumed a connection

between the chaining mechanism and the process of polysemy, or word sense extension. For instance, in both of his accounts on the extension of classifier systems and spatial prepositions, he described how polysemous extensions (e.g., how Dyirbal classifiers group ideas related to women, fire, and dangerous things (Lakoff, 1987), and how English *over* expresses a broad range of spatial configurations and metaphorical senses (Lakoff, 1987; Brugman, 1988)) might depend on image schematic transformations that are reflected through a process of chaining where one referent or sense links to another in complex chain-like structures. Recent computational work has extended these ideas in a formal setting and found that models of chaining—similar to those described in this chapter—can explain historical word sense extension in the English lexicon (Ramiro et al., 2018), although such models are far from perfect.

A caveat in both that study and Lakoff (1987)’s work is the under-specification of the diverse knowledge involved in word meaning extension and semantic change, which is clearly beyond the embedding-based semantic representation presented here. In this respect, whether or how the theory of chaining can explain the diverse range of semantic change in adjectives and other word classes remains an important open question.

### 6.3 Toward a general theory of word meaning extension

The lexicon is an innovative product of the mind, and here we have focused on examining one critical form of lexical innovation that involves word meaning extension. A general account of word meaning extension in natural language ought to explain how it functions at different temporal scales not restricted to a historical setting.

There are at least three levels at which word meaning extension can occur, summarized in decreasing temporal scales: 1) across languages, the relics of word meaning extension are reflected in the colexification and polysemy structures that are likely a result of language evolution through tens of thousands of years (François, 2008; Youn et al., 2016), e.g., how a single word form like *fire* can denote the senses of ‘physical fire’, ‘flame’, and ‘anger’; 2) within a language, word meaning extension can occur in language change during hundreds of years (Sweetser, 1991; Traugott and Dasher, 2001), e.g., how words like *mouse* originally referred to ‘a type of rodent’ later extended to express ‘a computer device’; 3) in child development typically within the first 2-3 years of life, children extend word meaning toward novel objects for which they lack the proper words in the form of overextension (Vygotsky, 1962; Clark, 1978; Rescorla, 1980), e.g., how children use *ball* to refer to ‘a balloon’. Characterizing the common mechanisms and knowledge underlying these phenomena will shed light on word meaning extension as a general strategy for making innovative use of a finite lexicon.

Recent studies have made initial progress toward this direction. For instance, Jr. and Xu (2019) developed a multi-modal semantic framework based on the exemplar model of chaining and showed that it predicts children’s overextension behavior in a variety of studies from the psychological and linguistic literature. Xu et al. (2020) showed that the frequency variation in cross-linguistic colexification, i.e., why certain senses are more commonly grouped (e.g., ‘fire’–‘flame’) under a single word form than others (e.g., ‘fire’–‘anger’) can be explained by a principle of cognitive economy, whereby senses that are frequently colexified across languages tend to be easily associable – an argument that is consistent with the chaining account presented here. However, critically lacking is demonstrating how the approaches and principles identified in language development and cross-linguistic settings can also explain historical semantic change attested in the world’s languages.

We believe that a general formal account of word meaning extension will involve three key ingredients related to the chaining processes discussed in this chapter: 1) algorithmic formulations such as the exemplar model described that capture the mechanisms of semantic chaining; 2) rich knowledge structures that support these mechanisms toward a diverse range of extensional strategies such as metonymy and metaphor; 3) external socio-cultural influences or events that provide the driving force for word meaning extension.

## 7 Conclusion

We have presented a computational approach to explore regularities in the historical composition of adjectives and nouns through probabilistic models of chaining. Our approach provides clues to the generative mechanisms that give rise to novel adjective usages over time, and we hope it will stimulate future work on the semantic representation and the interaction of cognitive and socio-cultural underpinnings of word meaning extension and semantic change.

## Acknowledgements

We thank Barend Beekhuizen, Lars Borin, Sammy Floyd, Adam Jatowt, Charles Kemp, and Nina Tahmasebi for helpful discussion and constructive comments. We also thank Amir Ahmad Habibi for sharing data and code. KG is supported by a Bell Graduate Scholarship and a Vector Scholarship in Artificial Intelligence. YX is funded through an NSERC Discovery Grant RGPIN-2018-05872 and a SSHRC Insight Grant #435190272.

## References

- M. Baroni and R. Zamparelli. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Cambridge, Massachusetts, 2010.
- G. Boleda, M. Baroni, N. Pham, and L. McNally. Intensionality was only alleged: On adjective-noun composition in distributional semantics. In *Proceedings of the seventh International Conference on Computational Semantics*, Potsdam, Brandenburg, 2013.
- C. M. Brugman. *The story of over: Polysemy, semantics, and the structure of the lexicon*. Garland, New York, New York, 1988.
- J. L. Bybee. Usage-based theory and exemplar representations of constructions. In T. Hoffmann and G. Trousdale, editors, *The Oxford handbook of construction grammar*. Oxford University Press, Oxford, Oxfordshire, 2013.
- J. L. Bybee, R. D. Perkins, and W. Pagliuca. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press, Chicago, Illinois, 1994.
- E. V. Clark. Strategies for communicating. *Child Development*, 49:953–959, 1978.
- A. François. *Semantic maps and the typology of colexication: Intertwining polysemy networks across languages*, pages 163–215. John Benjamins, Amsterdam, North Holland, 2008.
- K. Grewal and Y. Xu. Chaining and historical adjective extension. In *Proceedings of the forty-second Annual Meeting of the Cognitive Science Society*, Toronto, Ontario, 2020.
- A. A. Habibi, C. Kemp, and Y. Xu. Chaining and the growth of linguistic categories. *Cognition*, 202:104323, 2020.
- W. L. Hamilton, J. Leskovec, and D. Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the fifty-fourth Annual Meeting of the Association for Computational Linguistics*, Berlin, 2016.
- M. Hilpert. Chained metonymies in lexicon and grammar. *Aspects of meaning construction*, pages 77–98, 2007.
- R. F. P. Jr. and Y. Xu. Children’s overextension as communication by multimodal chaining. In *Proceedings of the forty-first Annual Meeting of the Cognitive Science Society*, Montréal, Québec, 2019.
- E. Keuleers. *Memory-based learning of inflectional morphology*. PhD thesis, Universiteit Antwerpen, 2008.

- G. Lakoff. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press, Chicago, Illinois, 1987.
- M. Lapata. A corpus-based account of regular polysemy: The case of context-sensitive adjectives. In *Proceedings of the second Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 63–70, 2001.
- M. Lapata, S. McDonald, and F. Keller. Determinants of adjective-noun plausibility. In *Proceedings of the ninth Conference of the European Chapter of the Association for Computational Linguistics*, Bergen, 1999.
- A. Lazaridou, G. Dinu, A. Liska, and M. Baroni. From visual attributes to adjectives through decompositional distributional semantics. *Transactions of the Association for Computational Linguistics*, 3:183–196, 2015.
- Y. Luo and Y. Xu. Stability in the temporal dynamics of word meanings. In *Proceedings of the fortieth Annual Conference of the Cognitive Science Society*, Madison, Wisconsin, 2018.
- B. C. Malt, S. A. Sloman, S. Gennari, M. Shi, and Y. Wang. Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, 40:230–262, 1999.
- J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, et al. Quantitative analysis of culture using millions of digitized books. *Science*, 331:176–182, 2011.
- T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, Stateline, Nevada, 2013.
- G. A. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*, 39(11): 39–41, 1995.
- T. Nagarajan and K. Grauman. Attributes as operators: Factorizing unseen attribute-object compositions. In *Proceedings of the fifteenth European Conference on Computer Vision*, Munich, Bavaria, 2018.
- R. M. Nosofsky. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115:39–57, 1986.
- J. B. Pierrehumbert. Exemplar dynamics: Word frequency, lenition and contrast. *Typological Studies in Language*, 45:137–158, 2001.
- C. Ramiro, M. Srinivasan, B. C. Malt, and Y. Xu. Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences*, 115:2323–2328, 2018.
- R. Ramsey. *An exemplar-theoretic account of word senses*. PhD thesis, Northumbria University, 2017.
- L. A. Rescorla. Overextension in early language development. *Journal of Child Language*, 7 (2):321–335, 1980.
- E. Rosch. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104:192–233, 1975.
- L. A. Schmidt, C. Kemp, and J. B. Tenenbaum. Nonsense and sensibility: Inferring unseen possibilities. In *Proceedings of the twenty-eighth Annual Conference of the Cognitive Science Society*, Vancouver, British Columbia, 2006.
- R. Skousen. *Analogical modeling of language*. Kluwer, Dordrecht, South Holland, 1989.

- S. A. Sloman, B. C. Malt, and A. Fridman. *Categorization versus similarity: The case of container names*, pages 73–86. Oxford University Press, New York, New York, 2001.
- J. Snell, K. Swersky, and R. S. Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, Long Beach, California, 2017.
- M. Steyvers and J. B. Tenenbaum. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29:41–78, 2005.
- E. Sweetser. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge University Press, Cambridge, Cambridgeshire, 1991.
- E. C. Traugott and R. B. Dasher. *Regularity in semantic change*. Cambridge University Press, Cambridge, Cambridgeshire, 2001.
- E. M. Vecchi, R. Zamparelli, and M. Baroni. Studying the recursive behaviour of adjectival modification with compositional distributional semantics. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Washington, 2013.
- E. M. Vecchi, M. Marelli, R. Zamparelli, and M. Baroni. Spicy adjectives and nominal donkeys: Capturing semantic deviance using compositionality in distributional spaces. *Cognitive Science*, 41(2):102–136, 2017.
- L. S. Vygotsky. *Language and thought*. MIT Press, Cambridge, Massachusetts, 1962.
- J. M. Williams. Synaesthetic adjectives: A possible law of semantic change. *Language*, 32: 461–78, 1976.
- K. Xu, J. L. Ba, R. Kiros, K. Cho, A. Courville, R. S. Z. Ruslan Salakhutdinov, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *Proceedings of the thirty-second International Conference on Machine Learning*, Lille, Hauts-de-France, 2015.
- Y. Xu, T. Regier, and B. C. Malt. Historical Semantic chaining and efficient communication: The case of container names. *Cognitive Science*, 40:2081–2094, 2016.
- Y. Xu, K. Duong, B. C. Malt, S. Jiang, and M. Srinivasan. Conceptual relations predict colexification across languages. *Cognition*, 201:104280, 2020.
- H. Youn, L. Sutton, E. Smith, C. Moore, J. F. Wilkins, I. Maddieson, W. Croft, and T. Bhattacharya. On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences*, 113:1766–1771, 2016.
- F. M. Zanzotto, I. Korkontzelos, F. Fallucchi, and S. Manandhar. Estimating linear models for compositional distributional semantics. In *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, 2010.

## Appendix A – Adjective sets

Here we present all adjectives used in our analysis, namely from the FRQ-200, RAND-200, and SYN-65 adjective sets. Adjectives with an asterisk (\*) are included in at least two of the three adjective sets. The first table gives the adjectives that constitute SYN-65, and we note two important details about this set. First, the set of synaesthetic adjectives proposed by Williams (1976) actually contains 64 unique adjectives as *light* is repeated. Second, the Google Books corpus ties all tokens to words in WordNet, and since *acrid*, *aspre*, and *tart* (all synaesthetic adjectives) are not WordNet adjectives, we could not reliably measure their uses through time. For this reason, we excluded these from SYN-65 and have 61 adjectives in total, listed in Table 3.

Table 3: List of 61 adjectives in SYN-65. Adjectives with an asterisk (\*) appear in at least one of FRQ-200 and RAND-200 as well.

SYN-65						
<i>acute</i>	<i>cloying</i>	<i>dulcet*</i>	<i>grave</i>	<i>light</i>	<i>quiet</i>	<i>sour</i>
<i>austere</i>	<i>coarse</i>	<i>dull</i>	<i>hard</i>	<i>little</i>	<i>rough</i>	<i>strident</i>
<i>big</i>	<i>cold*</i>	<i>eager</i>	<i>harsh</i>	<i>loud</i>	<i>shallow</i>	<i>sweet</i>
<i>bitter*</i>	<i>cool</i>	<i>empty</i>	<i>heavy</i>	<i>low</i>	<i>sharp</i>	<i>thick</i>
<i>bland</i>	<i>crisp</i>	<i>even</i>	<i>high</i>	<i>mellow</i>	<i>shrill</i>	<i>thin</i>
<i>bright</i>	<i>dark</i>	<i>faint</i>	<i>hollow</i>	<i>mild</i>	<i>small</i>	<i>vivid</i>
<i>brilliant*</i>	<i>deep</i>	<i>fat</i>	<i>hot</i>	<i>piquant</i>	<i>smart</i>	<i>warm*</i>
<i>brisk</i>	<i>dim</i>	<i>flat</i>	<i>keen</i>	<i>poignant</i>	<i>smooth</i>	
<i>clear</i>	<i>dry*</i>	<i>full</i>	<i>level</i>	<i>pungent</i>	<i>soft</i>	

Next, we present the FRQ-200 and RAND-200 adjective sets and the clusters we used for the analysis, listed in Table 4. Since these two sets draw adjectives from identical clusters, we present the two adjective sets so we can easily compare adjectives drawn from same cluster between the two sets.

Table 4: Lists of adjectives and clusters in FRQ-200 and RAND-200.

cluster 1 of 20		cluster 2 of 20		cluster 3 of 20	
FRQ-200	RAND-200	FRQ-200	RAND-200	FRQ-200	RAND-200
<i>casual</i>	<i>amiable</i>	<i>bare</i>	<i>contorted</i>	<i>sufficient</i>	<i>alterable</i>
<i>eccentric</i>	<i>chatty</i>	<i>curly</i>	<i>dainty</i>	<i>analogous</i>	<i>contemporaneous</i>
<i>energetic</i>	<i>communicative</i>	<i>eyed</i>	<i>furrowed</i>	<i>equal</i>	<i>reconcilable</i>
<i>entertaining</i>	<i>fiery</i>	<i>female</i>	<i>hale</i>	<i>calculable</i>	<i>chargeable</i>
<i>enthusiastic</i>	<i>fluent</i>	<i>feminine</i>	<i>horny*</i>	<i>receivable</i>	<i>distributive</i>
<i>forgiving</i>	<i>guileless</i>	<i>horny*</i>	<i>limber</i>	<i>derived</i>	<i>accessary</i>
<i>glib</i>	<i>lovable</i>	<i>male</i>	<i>sage</i>	<i>binding</i>	<i>lineal</i>
<i>intelligent</i>	<i>loyal</i>	<i>naked</i>	<i>skeletal*</i>	<i>indirect</i>	<i>allotted</i>
<i>passionate</i>	<i>patriotic</i>	<i>pale</i>	<i>smoky</i>	<i>undivided</i>	<i>noncommercial</i>
<i>polite</i>	<i>unorthodox</i>	<i>skeletal*</i>	<i>swaggering</i>	<i>eligible</i>	<i>classifiable</i>
cluster 4 of 20		cluster 5 of 20		cluster 6 of 20	
FRQ-200	RAND-200	FRQ-200	RAND-200	FRQ-200	RAND-200
<i>cold*</i>	<i>chilly</i>	<i>algebraic</i>	<i>binary</i>	<i>blind</i>	<i>intact</i>
<i>dense</i>	<i>cold*</i>	<i>conventional</i>	<i>biotic</i>	<i>impossible</i>	<i>irretrievable</i>
<i>dry*</i>	<i>drizzling</i>	<i>discrete</i>	<i>crystalline</i>	<i>incomplete</i>	<i>malfunctioning</i>
<i>eastern</i>	<i>encroaching</i>	<i>electrical</i>	<i>fusible</i>	<i>isolated</i>	<i>obscure</i>
<i>hardy</i>	<i>fertile</i>	<i>microscopic</i>	<i>geometric</i>	<i>pregnant</i>	<i>overlooked</i>
<i>northern</i>	<i>funicular</i>	<i>multicellular</i>	<i>interfacial</i>	<i>scarce</i>	<i>powerless</i>
<i>south</i>	<i>homeward</i>	<i>predictive</i>	<i>modular</i>	<i>silent</i>	<i>unmarked</i>
<i>tropical</i>	<i>littoral</i>	<i>rotational</i>	<i>perceptual</i>	<i>submerged</i>	<i>unstable</i>
<i>warm*</i>	<i>unincorporated</i>	<i>thermal</i>	<i>refrigerant</i>	<i>unknown</i>	<i>unstudied</i>
<i>western</i>	<i>watery</i>	<i>volcanic</i>	<i>stratified</i>	<i>unrelated</i>	<i>valueless</i>
cluster 7 of 20		cluster 8 of 20		cluster 9 of 20	
FRQ-200	RAND-200	FRQ-200	RAND-200	FRQ-200	RAND-200
<i>appropriate</i>	<i>complex</i>	<i>alien</i>	<i>antipodal</i>	<i>everyday</i>	<i>approaching</i>
<i>balanced</i>	<i>delighted</i>	<i>colonial</i>	<i>congruous</i>	<i>firm</i>	<i>descending</i>
<i>basic</i>	<i>foolproof</i>	<i>divine</i>	<i>dynastic</i>	<i>more</i>	<i>fiddling</i>
<i>better</i>	<i>grateful</i>	<i>heavenly</i>	<i>hierarchical</i>	<i>original</i>	<i>former</i>
<i>different</i>	<i>intensive</i>	<i>human</i>	<i>invariable</i>	<i>particular</i>	<i>intensifying</i>
<i>natural</i>	<i>knowledgeable</i>	<i>inner</i>	<i>overt</i>	<i>physical</i>	<i>probable</i>
<i>positive</i>	<i>livable</i>	<i>medieval</i>	<i>paschal</i>	<i>preliminary</i>	<i>rental</i>
<i>solid</i>	<i>realistic</i>	<i>modern</i>	<i>protestant</i>	<i>same</i>	<i>reverse</i>
<i>superior</i>	<i>structured</i>	<i>moral</i>	<i>recessive</i>	<i>several</i>	<i>sliding</i>
<i>sure</i>	<i>varied</i>	<i>philosophical</i>	<i>sacred</i>	<i>top</i>	<i>thirteenth</i>
cluster 10 of 20		cluster 11 of 20		cluster 12 of 20	
FRQ-200	RAND-200	FRQ-200	RAND-200	FRQ-200	RAND-200
<i>allergic</i>	<i>carcinogenic</i>	<i>black</i>	<i>ceramic</i>	<i>bent</i>	<i>hysterical</i>
<i>antibiotic</i>	<i>coagulate</i>	<i>circular</i>	<i>cyclopean</i>	<i>bourgeois</i>	<i>inattentive</i>
<i>artificial</i>	<i>colorless</i>	<i>concave</i>	<i>fireproof</i>	<i>corrupt</i>	<i>irreligious</i>
<i>dietary</i>	<i>milky</i>	<i>crimson</i>	<i>legible</i>	<i>disreputable</i>	<i>lunatic</i>
<i>fibrous</i>	<i>nonfat</i>	<i>distinctive</i>	<i>rectilinear</i>	<i>domineering</i>	<i>opportunist</i>
<i>liquid</i>	<i>pulpy</i>	<i>fluorescent</i>	<i>sleek</i>	<i>evil</i>	<i>parochial</i>
<i>mucous</i>	<i>scented</i>	<i>incised</i>	<i>tucked</i>	<i>fascist</i>	<i>possessive</i>
<i>powdery</i>	<i>spongy</i>	<i>red</i>	<i>umber</i>	<i>jugular</i>	<i>resentful</i>
<i>raw</i>	<i>steamed</i>	<i>tubular</i>	<i>unglazed</i>	<i>pious</i>	<i>uncongenial</i>
<i>synthetic</i>	<i>vanilla</i>	<i>white</i>	<i>Venetian</i>	<i>warlike</i>	<i>unengaged</i>

cluster 13 of 20		cluster 14 of 20		cluster 15 of 20	
FRQ-200	RAND-200	FRQ-200	RAND-200	FRQ-200	RAND-200
<i>bitter*</i>	<i>brokenhearted</i>	<i>affected</i>	<i>bottomed</i>	<i>abusive</i>	<i>appalling</i>
<i>debilitating</i>	<i>confused</i>	<i>buried</i>	<i>credited</i>	<i>deplorable</i>	<i>bias</i>
<i>emotional</i>	<i>delirious</i>	<i>distributed</i>	<i>jammed</i>	<i>exaggerated</i>	<i>capricious</i>
<i>hopeless</i>	<i>disturbed</i>	<i>given</i>	<i>owned</i>	<i>excessive</i>	<i>exorbitant</i>
<i>odd*</i>	<i>odd*</i>	<i>left</i>	<i>rose</i>	<i>illegal</i>	<i>hostile</i>
<i>poor</i>	<i>patchy</i>	<i>marked</i>	<i>scattered*</i>	<i>simplistic</i>	<i>imprecise</i>
<i>troubled</i>	<i>regretful</i>	<i>modified</i>	<i>settled</i>	<i>undue</i>	<i>inelegant</i>
<i>unhappy*</i>	<i>thirsty</i>	<i>scattered*</i>	<i>shattered</i>	<i>unintentional</i>	<i>innocuous</i>
<i>weird</i>	<i>unhappy*</i>	<i>used</i>	<i>surrounded</i>	<i>unproductive</i>	<i>unbalanced</i>
<i>worst</i>	<i>untidy</i>	<i>worn</i>	<i>sworn</i>	<i>wrong</i>	<i>unsound</i>
cluster 16 of 20		cluster 17 of 20		cluster 18 of 20	
FRQ-200	RAND-200	FRQ-200	RAND-200	FRQ-200	RAND-200
<i>adrenal</i>	<i>cesarean</i>	<i>American</i>	<i>Arabian</i>	<i>brilliant*</i>	<i>adored</i>
<i>alveolar</i>	<i>endoscopic</i>	<i>Asian</i>	<i>Catalan</i>	<i>conspicuous</i>	<i>commanding</i>
<i>bivariate</i>	<i>hemorrhagic</i>	<i>Christian</i>	<i>Chinese</i>	<i>ecstatic</i>	<i>fantastic</i>
<i>cardiovascular</i>	<i>hyoid</i>	<i>Dutch*</i>	<i>Cornish</i>	<i>extraordinary</i>	<i>favorite</i>
<i>clinical</i>	<i>intervertebral</i>	<i>English</i>	<i>Dutch*</i>	<i>fitting</i>	<i>gallant</i>
<i>diagnostic</i>	<i>lobular</i>	<i>European</i>	<i>Haitian</i>	<i>great</i>	<i>halcyon</i>
<i>neural</i>	<i>monovalent</i>	<i>French</i>	<i>Hungarian</i>	<i>incomparable</i>	<i>loved</i>
<i>peritoneal</i>	<i>normotensive</i>	<i>Roman</i>	<i>Kurdish</i>	<i>perfect</i>	<i>superb</i>
<i>spinal</i>	<i>valved</i>	<i>Serbian</i>	<i>Taiwanese</i>	<i>singular</i>	<i>tragic</i>
<i>ulcerative</i>	<i>vesicular</i>	<i>Spanish</i>	<i>Thai</i>	<i>startling</i>	<i>undefeated</i>
cluster 19 of 20		cluster 20 of 20			
FRQ-200	RAND-200	FRQ-200	RAND-200		
<i>budgetary</i>	<i>agrarian</i>	<i>aesthetic</i>	<i>clarion</i>		
<i>civil</i>	<i>catechetical</i>	<i>artistic</i>	<i>contemporary</i>		
<i>criminal</i>	<i>clandestine</i>	<i>classical</i>	<i>darkling</i>		
<i>marital</i>	<i>constitutional</i>	<i>clever</i>	<i>dulcet*</i>		
<i>mental</i>	<i>curricular</i>	<i>colloquial</i>	<i>earthy</i>		
<i>national</i>	<i>hourly</i>	<i>dreamy</i>	<i>falsetto</i>		
<i>nuclear</i>	<i>intramural</i>	<i>hilarious</i>	<i>longhand</i>		
<i>parental</i>	<i>qualitative</i>	<i>intimate</i>	<i>ponderous</i>		
<i>regional</i>	<i>recreational</i>	<i>narrative</i>	<i>soothing</i>		
<i>social</i>	<i>sectional</i>	<i>rhetorical</i>	<i>wry</i>		

## Appendix B – Temporal trends in model precision

As discussed in the main text, the model precision generally decreases across all models with time. As Figure 7 shows, the average number of nouns to predict in each decade decreases with time. This trend applies to both sets of true positives: only adjectives that first co-occur with a given noun  $n^*$  in decade  $t + \Delta$ , and also in any future decade. Consequently, the precision falls systematically in later decades because there are fewer novel pairings to predict in the data.

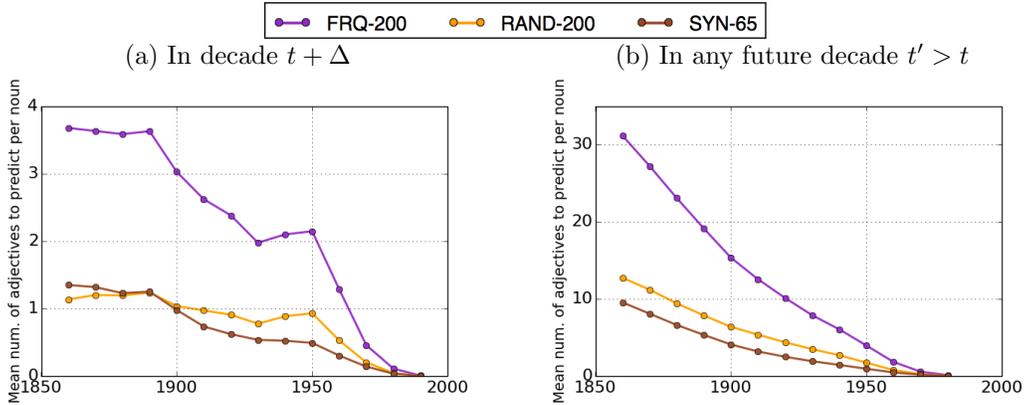


Figure 7: The average number of novel adjective-noun pairs remaining for each model to predict across all times and adjective sets. This value is computed across all nouns for which a predictive model makes adjective prediction.

## Appendix C – Semantically changing and stable adjectives

Table 5: Lists of most and least changed adjectives from the FRQ-200, RAND-200, and SYN-65 sets, along with their top semantic neighbours during initial (1850s) and terminal (1990s) periods of investigation.

FRQ-200	
Least Changed	Most Changed
1. <i>eccentric</i> 1850s: <i>versatile, droll, impulsive</i> 1990s: <i>incoherent, perverse, exquisite</i>	1. <i>classical</i> 1850s: <i>theological, modern, greek</i> 1990s: <i>greek, traditional, contemporary</i>
2. <i>casual</i> 1850s: <i>occasional, trivial, careless</i> 1990s: <i>careless, informal, friendly</i>	2. <i>rhetorical</i> 1850s: <i>idiomatic, didactic, fanciful</i> 1990s: <i>poetic, epistolary, grammatical</i>
3. <i>polite</i> 1850s: <i>affable, hospitable, elegant</i> 1990s: <i>respectful, friendly, agreeable</i>	3. <i>colloquial</i> 1850s: <i>imaginative, analytic, bewitching</i> 1990s: <i>epistolary, poetic, idiomatic</i>
4. <i>intelligent</i> 1850s: <i>honest, rational, inquisitive</i> 1990s: <i>clever, energetic, minded</i>	4. <i>narrative</i> 1850s: <i>detailed, circumstantial, brief</i> 1990s: <i>autobiographical, biblical, historical</i>
5. <i>enthusiastic</i> 1850s: <i>irrepressible, impulsive, passionate</i> 1990s: <i>ardent, sincere, generous</i>	5. <i>artistic</i> 1850s: <i>scientific, architectural, literary</i> 1990s: <i>intellectual, musical, poetic</i>

RAND-200	
Least Changed	Most Changed
1. <i>fluent</i> 1850s: <i>versatile, idiomatic, sprightly</i> 1990s: <i>spoken, speaking, Arabic</i>	1. <i>contemporary</i> 1850s: <i>recorded, voluminous, anonymous</i> 1990s: <i>literary, historical, classical</i>
2. <i>amiable</i> 1850s: <i>humane, affable, estimable</i> 1990s: <i>dignified, virtuous, pleasing</i>	2. <i>earthy</i> 1850s: <i>alkaline, gelatinous, nitrogenous</i> 1990s: <i>ceremonious, ravaging, disused</i>
3. <i>patriotic</i> 1850s: <i>loyal, disinterested, enlightened</i> 1990s: <i>democratic, civic, loyal</i>	3. <i>soothing</i> 1850s: <i>melancholy, sweet, sympathetic</i> 1990s: <i>calm, sweet, shrill</i>
4. <i>fiery</i> 1850s: <i>fierce, resistless, malign</i> 1990s: <i>mutinous, treacherous, fierce</i>	4. <i>ponderous</i> 1850s: <i>huge, cased, jingling</i> 1990s: <i>glistening, ethereal, noiseless</i>
5. <i>communicative</i> 1850s: <i>sociable, choleric, affable</i> 1990s: <i>symbolic, verbal, functional</i>	5. <i>clandestine</i> 1850s: <i>nefarious, illicit, adulterous</i> 1990s: <i>disfigured, patrician, sedate</i>

SYN-65	
Least Changed	Most Changed
1. <i>bitter</i> 1850s: <i>astringent, sweet, poignant</i>	1. <i>shrill</i> 1850s: <i>blithe, deafening, inaudible</i>

- |                                               |                                             |
|-----------------------------------------------|---------------------------------------------|
| 1990s: <i>sour, harsh, intense</i>            | 1990s: <i>pitched, startled, muffled</i>    |
| 2. <i>bland</i>                               | 2. <i>small</i>                             |
| 1850s: <i>mild, unobtrusive, affable</i>      | 1850s: <i>smaller, size, sized</i>          |
| 1990s: <i>unconverted, unadorned, affable</i> | 1990s: <i>sized, smaller, insignificant</i> |
| 3. <i>coarse</i>                              | 3. <i>mellow</i>                            |
| 1850s: <i>dirty, threadbare, boned</i>        | 1850s: <i>lustrous, chilly, balmy</i>       |
| 1990s: <i>thin, fine, stiff</i>               | 1990s: <i>perfumed, fragrant, sportive</i>  |
| 4. <i>cold</i>                                | 4. <i>austere</i>                           |
| 1850s: <i>clammy, wet, hot</i>                | 1850s: <i>unsocial, disdainful, rigid</i>   |
| 1990s: <i>warm, damp, windy</i>               | 1990s: <i>matchless, apposite, erudite</i>  |
| 5. <i>cool</i>                                | 5. <i>pungent</i>                           |
| 1850s: <i>calm, chilly, warm</i>              | 1850s: <i>juicy, ductile, astringent</i>    |
| 1990s: <i>damp, hot, dry</i>                  | 1990s: <i>mown, fresh, colorless</i>        |
-