The Dynamics of Semantic Change: A Corpus-Based Analysis

Mohamed Amine Boukhaled$^{(\boxtimes)}$, Benjamin Fagard, and Thierry Poibeau

Laboratoire Langues, Textes, Traitements Informatique, Cognition,
LATTICE, CNRS, ENS & Université Paris 3; PSL & USPC, Paris, France
{amine.boukhaled,benjamin.fagard,thierry.poibeau}@ens.fr

Abstract. In this contribution, we report on a computational corpus-based study to analyse the semantic evolution of words over time. Though semantic change is complex and not well suited to analytical manipulation, we believe that computational modelling is a crucial tool to study this phenomenon. This study consists of two parts. In the first one, our aim is to capture the systemic change of word meanings in an empirical model that is also predictive, making it falsifiable. In order to illustrate the significance of this kind of empirical model, we then conducted an experimental evaluation using the Google Books N-Gram corpus. The results show that the model is effective in capturing semantic change and can achieve a high degree of accuracy on predicting words' distributional semantics. In the second part, we look at the degree to which the S-curve model, which is generally used to describe the quantitative property associated with linguistic changes, applies in the case of lexical semantic change. We use an automatic procedure to empirically extract words that have known the biggest semantic shifts in the past two centuries from the Google Books N-gram corpus. Then, we investigate the significance of the S-curve pattern in their frequency evolution. The results suggest that the S-curve pattern has indeed some generic character, especially in the case of frequency rises related to semantic expansions.

Keywords: Semantic change · Diachronic word embedding · Recurrent neural networks · Computational semantics · S-curve model

1 Introduction

The availability of very large textual corpora spanning several centuries has recently made it possible to observe empirically the evolution of language over time. This observation can be targeted toward a few isolated words or a specific linguistic phenomenon, but it can also be interesting to combine these specific studies with the search for more general laws of language change and evolution. Semantic change, on which we shall focus in this contribution, includes all changes affecting the meaning of lexical items over time. For example, the word *awful* has drastically changed in meaning, moving away from a rather positive connotation, as an equivalent of *impressive* or *majestic*, at the beginning of the nineteenth century, and toward a negative one, as an equivalent of *disgusting* and *messy* nowadays [1]. It has been established that there are some systemic

regularities that direct the semantic shifts of words meanings. Not all words exhibit the same degree and speed of semantic change. Some words (or word categories) might be more resistant than others to the phenomenon of semantic change, as proposed by Dubossarsky et al. [2]. Various hypotheses have been proposed in the literature to explain such regularities in semantic change from a linguistic point of view [3].

One of the main challenges facing researchers studying the phenomenon of semantic change is its formidable complexity. It seems impossible to grasp all details and factors involved in this type of change, its abstract nature making it analytically intractable. However, computational models have no difficulties in handling complexity, and can therefore be used as means to make the study of semantic changes more accessible. Computational modelling of language change is a relatively new discipline, which includes early works that aimed at characterising the evolution through statistical and mathematical modelling [4, 5] and more recent and advanced works involving artificial intelligence, robotics and large-scale computer simulations [6].

In this context, the computational study of text temporality in general and semantic change in particular has become an active research topic, especially with the emergence of new and more effective methods of numerical word representations. The interest of taking into account the temporal dimension and the diachronic nature of meaning change as a research direction has been effectively demonstrated in several studies. It makes it possible to model the dynamics of semantic change [7], to analyse trajectories of meaning change for an entire lexicon [8], to model temporal word analogy or relatedness [9, 10], to capture the dynamics of semantic relations [11], and even to spell out specific laws of semantic change, among which:

- The Law of Conformity, according to which frequency is negatively correlated with semantic change [12].
- The Law of Innovation, according to which polysemy is positively correlated with semantic change [12].
- The Law of Prototypicality, according to which prototypicality is negatively correlated with semantic change [2].

In other studies on language change, researchers have been more interested in analysing the quantitative patterns associated with the propagation of changes. It has been shown that the diffusion of linguistic changes over time can be commonly presented as a sigmoidal curve (slow start/latency, accelerating period followed by a slow end/saturation) [13]. In this matter, the so-called S-curve model is generally used to describe the quantitative properties of frequency profiles. However, there have been only few computational studies in the literature to attest this quantitative pattern, and to the best of our knowledge, this claim is yet to be supported by case studies in the case of lexical semantic change as well.

In this work, we address the question of semantic change from a computational point of view. We conduct a computational corpus-based study that consists of two parts. In the first part, our aim is to capture the systemic change of word meanings in an empirical model that is also predictive, contrary to most previous approaches that meant to reproduce empirical observations. In the second part, we try to analyse the degree to which the S-curve model applies in the case of phenomena of lexical semantic change.

Both parts of our study were conducted using a large-scale diachronic corpus, namely the Google Books N-gram corpus.

The rest of the chapter is organised as follows: Sect. 2 presents the concept of diachronic word embedding and describes how this technique can be used to empirically quantify the degree of semantic change. Section 3 presents the first part of our study consisting of modelling and predicting semantic change using diachronic word embedding. In Sect. 4, we report on the second part of our contribution concerned with analysing the quantitative behaviour of some cases of semantic change. Finally, Sect. 5 concludes the chapter.

2 Empirical Assessment of the Semantic Change Using Diachronic Word Embedding

2.1 Diachronic Word Embedding

To represent computationally the meaning of words over time-periods, it is necessary first to extract the embedded projections of these words in a continuous vector space according to their contextual relationships [14]. Various methods can be used to obtain such vectors, such as Latent Semantic Analysis [15] and Latent Direchlet Allocation [16]. However, more recent and advanced techniques such as word2vec [17] and GloVe [18], known commonly as word embedding techniques, seem capable of better representing the semantic properties and the contextual meaning of words compared to traditional methods. Indeed, word embedding techniques have established themselves as an important step in the processing pipeline of natural languages.

The word2vec algorithm is one of the most frequently used techniques to construct word embeddings, with a huge impact in the field. It consists in training a simple neural network with a single hidden layer to perform a certain task (see Fig. 1). Training is achieved through stochastic gradient descent and back-propagation.

In the case of the skip-gram with negative sampling (SGNS) variant of the algorithm [17], the learning task is as follows: Given a specific word in the middle of a sentence, the model uses this current word to predict the surrounding window of context words. The words are in fact projected from a discrete space of V dimensions (where V is the vocabulary size) onto a lower dimensional vector space using the neural network. The goal is not to use the network afterward. Instead, it is just to learn the weights of the hidden layer. These weights constitute actually the word embedding vectors. Despite its simplicity, the word2vec algorithm, given an appropriate amount of training text, is highly effective in capturing the contextual semantics of words.

Such word embedding techniques can be tweaked to work in a diachronic perspective. The method consists first in training and constructing embeddings for each time-period, then in aligning them temporally, so as to finally use them as a means to track semantic change over time and thus to identify the words that have known the biggest semantic change in the corpus.

In our case, we used pre-trained diachronic word embeddings constructed on the basis of time-periods measured in decades from 1800 to 1990 [12]. The training text used to produce these word embeddings is derived from the Google Books N-gram

Input Word   Word Embedding   Output Word
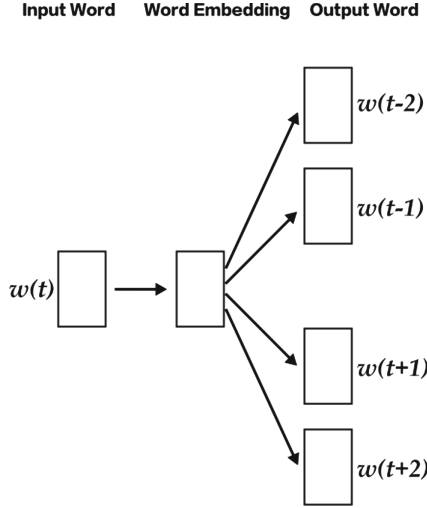
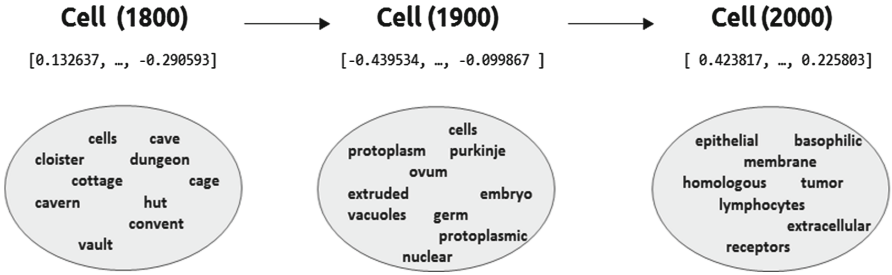**Fig. 1.** Architecture of the Skip-gram Model [17].

datasets [19] which contain large amounts of historical texts in many languages (N-Grams from approximately 8 million books, roughly 6% of all books published at that time). Each word in the corpus appearing from 1800 to 1999 is represented by a set of twenty continuous 300-th dimensional vectors; one vector for each decade.

2.2   Quantifying the Degree of Semantic Change

From a technical point of view, the degree of semantic change can be computationally examined with the help of mainly two different measures using diachronic word embeddings. The first one, known as the global measure, simply consists in computing the cosine distance between a given word's vectors from two consecutive decades t and $t + 1$. The bigger the distance, the higher the semantic change [8]. The second measure, which we chose to use in our work, is known as the local neighbourhood measure, recommended by Hamilton et al. [20]. It consists in evaluating the amount of semantic change of a word based on how much its corresponding semantic neighbours have changed between two consecutive decades, as illustrated in Fig. 2. To do so, we first extract for each word x_i, with its corresponding embedding vector w_i, the set of k nearest neighbours, denoted by $N_k(x_i)$, according to cosine-similarity for both consecutive decades t and $t+1$. Then, in order to measure the change which took place between these two decades, we compute a second-order similarity vector for $x_i^{(t)}$ from these neighbour sets. This second-order vector, denoted by $s_i^{(t)}$, contains the cosine similarity of w_i and the vectors of all x_i's nearest semantic neighbours in the time-periods t and $t + 1$, with entries defined as:

$$S^{(t)}(j) = cosine\_sim\left(w_i^{(t)}, w_j^{(t)}\right) \forall x_j \in N_k\left(x_i^{(t)}\right) \cup N_k(x_i^{(t+1)}) \tag{1}$$

An analogous vector for $x_i^{(t+1)}$ is similarly computed as well.

Cell (1800) $\longrightarrow$ **Cell (1900)** $\longrightarrow$ **Cell (2000)**

[0.132637, …, -0.290593]   [-0.439534, …, -0.099867]   [0.423817, …, 0.225803]

cells   cave
cloister   dungeon
cottage   cage
cavern   hut
convent
vault

cells
protoplasm   purkinje
ovum
extruded   embryo
vacuoles   germ
protoplasmic
nuclear

epithelial   basophilic
membrane
homologous   tumor
lymphocytes
extracellular
receptors

Fig. 2. Visualisation of the semantic change in the English word "*cell*" using diachronic word embedding. In the early 19th century the word *cell* was typically used to refer to a prison cell, hence the frequency of *cage* and *dungeon* in the context of *cell* in 1800, whereas in the late 19th century its meaning changed as it came to be frequently used in a scientific context, referring to a microscopic part of a living being (see *protoplasm*, *ovum*, etc. in the 1900 context).

Finally, we compute the local neighbourhood distance that measures the extent to which x_i's similarity with its nearest neighbours has changed as:

$$d(x_i^{(t)}, x_i^{(t+1)}) = cosine\_dist\left(s_i^{(t)}, s_i^{(t+1)}\right) \qquad (2)$$

Hamilton et al. [20] have found that the local neighbourhood measure is more effective in capturing specific cultural and semantic shifts than the global measure, while being less sensitive to other types of change such as grammaticalization.

3   Modelling the Semantic Change Dynamics Using Diachronic Word Embedding

In this first part of our study, we aim to capture the systemic change of words meanings in an empirical model that can also predict such changes, making it falsifiable. Our goal is thus to define a model capable of learning how the meanings of words have changed over time, and then use this model to predict how these meanings may evolve. This can then be checked against the actual meaning change which can be assessed with the same corpus. We propose a model that consists of two components:

1. *Diachronic word embeddings* to represent the meanings of words over time as the data component of the model as described in the previous section.
2. *A recurrent neural network* to learn and predict the temporal evolution patterns of these data.

The idea behind our model is to train Long Short Term Memory units (LSTMs) Recurrent Neural Network (RNN) on word embeddings corresponding to given time-periods (measured in decades) and try to predict the word embeddings of the following decade. We then evaluate the model using the diachronic embeddings derived from the English Google Books N-Gram corpus. The next two subsections describe more thoroughly the architecture of the recurrent neural network used in our work and the experimental evaluation respectively.

3.1   Predicting Semantic Change with Recurrent Neural Networks

As we are interested in predicting a continuous vector of d dimensions representing a word's contextual meaning in a given decade, this task is considered to be a regression problem (by opposition to a classification problem, where the task is to predict a discrete class). Many algorithms have been proposed in the literature to deal with this kind of temporal pattern recognition problem, such as Hidden Markov Models [21] and Conditional Random Fields [22].

In this work, we propose to use a recurrent neural network with a many-to-one LSTMs architecture to address this problem. RNNs are a powerful class of artificial neural networks designed to recognise dynamic temporal behaviour in sequences of data such as textual data [23]. RNNs are distinguished from feed-forward networks by the feedback loop connected to their past states. In this feedback loop, the hidden state h_t at time step t is a function F of the input at the same time step x_t modified by a weight matrix W, added to the hidden state of the previous time step h_{t-1} multiplied by its own transition matrix U as in Eq. (3):

$$h_t = F(Wx_t + Uh_{t-1}) \tag{3}$$

More specifically, we used a LSTMs architecture [24] (see Fig. 3) which is a variety of RNNs designed to deal effectively with the problem of vanishing gradient that RNNs suffer from while training [25].
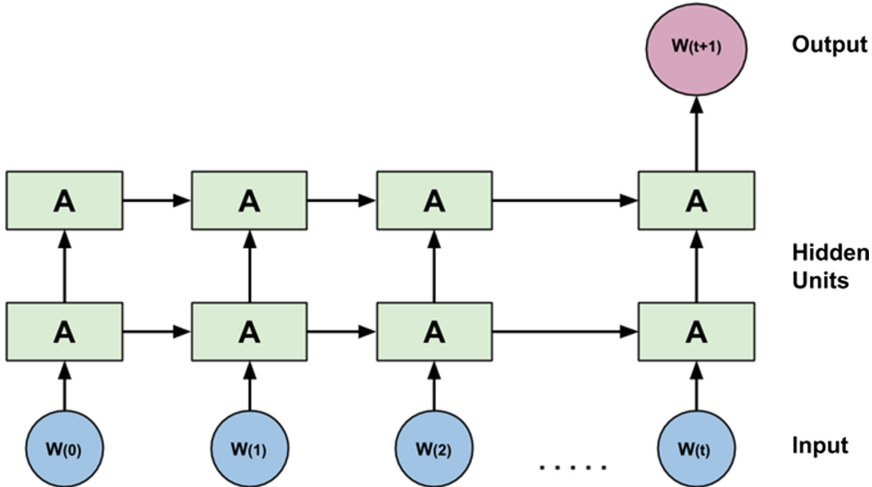
Fig. 3. The many-to-one LSTMs architecture used in our work to predict word embedding vectors. For each word in the vocabulary, the network is trained on diachronic word embedding vectors of time-periods $(1, \ldots, t)$ as input and tries to predict the embedding vector for time $t + 1$ as output [7].

Problem Formulation. In this contribution, we use the same mathematical formulation used in [7]. Let us consider a vocabulary V_n consisting in the top-n most frequent words of the corpus and $W(t) \in R^{d*n}$ to be the matrix of word embeddings at time step t.

The LSTMs network is trained on word embeddings over time-period $(1, \ldots, t)$ as input and asked to predict word embeddings $\widehat{W}(t+1)$ of time $t+1$ as output. The predicted embedding $\widehat{W}(t+1)$ is then compared to the ground truth word embedding $W(t+1)$ in order to assess the prediction accuracy. Predicting a continuous 300-th dimensional vector representing a word's contextual meaning is thus, as indicated above, formulated as a regression problem. Traditionally, researchers use mean-squared-error or mean absolute-error cost functions to assess the performance of regression algorithms. However, in our case, such cost functions would not be adapted, as they provide us with an overview (i.e., numerical value) of how the model is performing but little detail on its prediction accuracy. To have a more precise assessment of the prediction accuracy, we need to be able to say whether the prediction, for each word taken individually, is correct or not. Overall prediction accuracy is then computed. To do so, we proceed as follows:

Given the vocabulary V_n constituted from the top-n most frequent words and the matrix $W(t)$ of word embeddings at decade t, let us consider a word $x_i \in V_n$ and $\widehat{w}_i(t+1)$ its predicted word-embedding at decade $t+1$. Though it is virtually impossible to predict exactly the same ground truth vector $w_i(t+1)$ for this decade, as we are working on a continuous 300-th dimensional space, the accuracy of the predicted vector $\widehat{w}_i(t+1)$ can be assessed by extracting the words that are closest semantically, based on cosine-similarity measure. If the word x_i is actually the nearest semantic neighbour to the predicted vector $\widehat{w}_i(t+1)$, then it is considered to be a correct prediction. Otherwise, it is considered to be a false prediction.

3.2 Experimental Evaluation

In our experiment, we used word embeddings of all decades from 1800 to 1990 as input for the network, and we trained it to predict the embeddings of the 1990–1999 decade as output. We then conducted two types of evaluations. The first one consisted in reconstructing the overall large-scale evolution of the prediction accuracy on most frequents words from the corpus. The second one consisted in evaluating the prediction accuracy for a handful of word forms that have known considerable semantic shifts in the past two centuries. In what follows, we describe the experimental protocol and the results for both evaluations.

Overall Evaluation. In this first part of the evaluation, we experimented with different vocabulary sizes (top-1,000, 5,000, 10,000 and 20,000 words respectively, consisting of the most frequent words as computed from their average frequency over the entire historical time-periods). The experimental settings can help us evaluate the effect of a word's frequency on the degree of semantic change, and hence on prediction accuracy. For each experiment, and in order to get a reasonable estimate of the expected prediction performance, we used a 10-fold cross-validation method. Each time, we used 90% of the words for training the model, and the remaining 10% for testing its prediction accuracy. The training and testing process was repeated 10 times. The overall prediction accuracy is taken as the average performance over these 10 runs.

The results of measuring the prediction accuracy in our experimental evaluation are summarised in Table 1. They show that the model can be highly effective in capturing

semantic change and can achieve a high accuracy when predicting words' distributional semantics. For example, the model was able to achieve 71% accuracy trained and tested exclusively on embeddings coming from the 10,000 most frequent words of the corpus.

Table 1. Results of prediction accuracy measured for different vocabulary sizes [7].

Vocabulary size	Acc.
1,000	91.7%
5,000	86.1%
10,000	71.4%
20,000	52.2%

The results also show a better performance when using a smaller vocabulary size, containing only the most frequent words. This is due to the fact that frequent words are repeated a sufficient number of times for the embedding algorithm to represent them accurately, and therefore to have a better distinction of the semantic change pattern which those embeddings may contain, which in turn can lead the RNN model to better capture this semantic change pattern and yield a more accurate prediction. Indeed, having a large corpus is essential to enable the models to learn a better word representation. These results are also in line with previous works claiming that frequency plays an important role in the semantic change process. For instance, Hamilton et al. [12] have shown that frequent words tend to be more resistant to semantic change (statistical Law of Conformity).

Table 2. Neighbouring words according to cosine similarity (based on word2vec embeddings) for sample words that have known important semantic shifts in the past two centuries [7].

Word	Neighbours in 1800s	Neighbours in 1990s
Circuit	Habitable, district, lanes, range, area, outer, globe, traverse	Circuits, appeals, amplifier, voltage, transistor, capacitor, appellate, court, resistor, district
Signal	Commodore, hoisted, victory, tack, admiral, commemoration, victories, chace, flag, announce	Signals, modulated, amplitude, input, noise, modulation, transmitter, analog, waveform, transduction
Array	Banners, spears, shields, ensigns, ranged, pikes, trumpets, banner, standards	Arrays, variety, range, integers, integer, byte, wide, pointer, formats, pointers
Mail	Waistcoat, boots, shirt, gloves, breeches, velvet, pistols, shoe, helmet, spurs	Mailing, email, send, internet, telephone, sending, fax, messages, mails, postage

Case Studies. We further examined the prediction accuracy on a handful of words. Based on the automatic procedure to assess the degree of semantic change as described

in Sect. 2, we automatically extracted from the Google Books N-Gram Corpus the top-100 words that have known the most important semantic shifts in the past two centuries. We noticed that these words correspond mostly to cases that are correlated with important cultural shifts, which makes them a harder challenge for the prediction model compared to datasets used earlier in the overall evaluation. Table 2 presents sample words that gained new meanings due to their evolution towards uses in scientific and technological contexts. The model was able to correctly predict the semantic evolution of 41% of the studied cases, including words that have known an important and attested semantic change in the last two centuries such as the word *cell*. Moreover, a large portion of the false predictions corresponds to borderline cases for which the model has a tendency to predict vectors that are closer to much more frequent words, occurring in the same semantic context in the corpus, such as predicting a vector closer to the (emerging but more frequent) word *optimistic* for the (declining) word *sanguine*. The word *sanguine* comes from Old French *sanguin* (itself from Latin *sanguineus*, on *sanguis* 'blood'). It originally means 'blood-red' (14th c., Merriam-Webster's), and by extension 'hopeful, optimistic' (15th c., ibid.). In our corpus examples from the early 18th c., it is already used with the meaning 'optimistic', as in "My lords, I am sanguine enough to believe that this country has in a great measure got over its difficulties" (*Speech of the Earl of Liverpool*, 1820: 31) and "she is sanguine enough to expect that her various Novelties will meet the approbation of all Ladies of taste" (*La belle assemblee*, vol. XV, April 1st, 1817). But Fig. 4 shows that its frequency in the 19th and 20th c. has dropped steadily, while *optimistic* has seen its frequency rise sharply. Thus, the pair *sanguine/optimistic* seems to be a good example of lexical replacement, which explains our model's prediction.
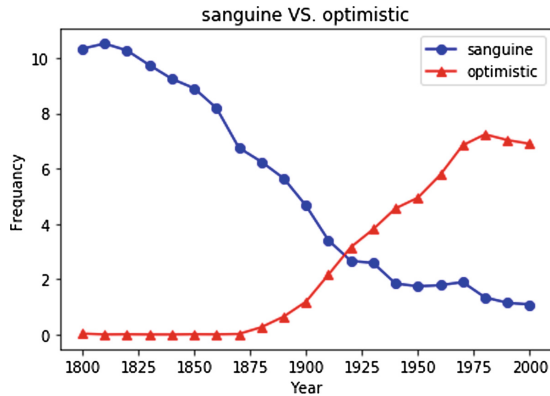
Fig. 4. Frequency profiles of *sanguine* and *optimistic* in Google Books N-Gram Corpus measured in millions [7].

Thus, among other benefits for historical linguists, our method makes it possible to identify the semantic replacement of one word by another in a specific context.

Despite being effective in predicting the semantic evolution of words, some difficulties remain regarding our method. For instance, our model works best for the most frequent words, i.e., according to the Law of Conformity, those with the least semantic

evolution. One could thus wonder whether the words for which the model correctly predicts the semantic are not simply those which display little or no semantic change. The examples given in these case studies show that this is at least not always the case, but a more systematic investigation of individual cases is needed in order to get a clear picture. Another way to answer this question would be to explore more finely the effect that both polysemy and word frequency may have on our results, especially on the word representation part of our model. These two factors have been shown to play an important role in the semantic change, and their effects need to be studied and formalized more explicitly. Exploring more advanced and semantic-oriented word embedding techniques, referred to as sense embeddings, such as SENSEMBED [26], could help make the model less sensitive to those factors. In the next section describing the second part of our contribution, we will take a closer look at the frequency aspect of that matter by investigating the frequency pattern associated with lexical semantic change.

4   Investigating the Frequency Pattern Associated with Lexical Semantic Change

In this second part of our study, our aim is to analyse the degree to which the S-curve model applies in the case of phenomena of lexical semantic change. We investigate the significance of this pattern through an empirical observation using the same large-scale diachronic corpus as before, namely Google Books N-gram corpus. The rest of the section is organized as follows: Subsect. 4.1 presents a brief review of the S-curve pattern in linguistics. Subsect. 4.2 describes how to automatically extract it from corpus data. In Subsect. 4.3, we describe our experimental settings and present the results.

4.1   The S-curve Pattern in Linguistics

S-curved functions have been used to analyse the diffusion of changes and innovations in various scientific domains [27]. They describe change as resulting from the fact that different parts of a given population successively adopt a new variant, which increases its frequency of use over time. This frequency subsequently reaches a saturation point at which the adoption of the variant stagnates.

The S-curve has been used starting from the early 1950s to model phenomena of language change [28]. It supposes that language change occurs commonly according to an S-curve, which means that the frequency rise associated with the new change variant should obey a sigmoid function, or any similar function which has approximately the same shape. The main reference in the literature on such matter is the work of Kroch [5]. Several other works follow [29–31]. Blythe and Croft [32] have performed a detailed survey on S-shaped patterns in linguistics totalising about 40 cases of changes. Feltgen et al. [33] have provided a statistical survey on frequency patterns associated with about 400 cases of functional semantic change (grammaticalization) in French. This investigation has shown that 70% of the studied cases display at least one S-curve increase of frequency in the course of their adoption. Apart from these studies and despite having an important part to play in the understanding of the underlying mechanisms of language evolution, it appears that large-scale quantitative studies of the statistical properties of

language change were left aside, limited mostly by both the quantity and the quality of the available historical linguistic data.

In semantic change, and more precisely in the case of semantic expansion, it is expected to observe an increase in frequency for the word whose meaning changes. In such cases, the generic character of the S-curve as a pattern associated with meaning change can be described by the following three steps:

1. The first occurrences of the word linked to the new meaning start to appear in the corpus and the frequency slightly increases to a certain low value. The word linked to the same new meaning continues to show in the corpus, but not so much as to reach a momentum. The frequency may thus remain constant for a certain period of time.
2. Following its adoption, the word carrying the new meaning is used in a broader number of contexts by an increasing portion of the population. The frequency rises rapidly.
3. At a certain point, the rise in frequency reaches a saturation point. Saturation occurs either because there is a limited number of new 'adopters' or because there is a limited number of compatible linguistic contexts in which the new meaning can be used [33].

In the following subsection, we describe how this pattern can be mathematically extracted from diachronic textual data.

4.2 Extraction of S-curve Patterns from a Diachronic Corpus

In order to identify and extract S-curve patterns in diachronic data, we used a similar procedure to the one described in [13]. The process is as follows:

1. For each word in the investigation dataset, we first extract its frequency x_i in each time period i of the diachronic corpus (in our case, we consider each year as one time period). These frequencies are smoothed by computing a moving average over the past five years.
2. For each frequency profile, we identify the two years i_{start} and i_{end} marking respectively the beginning and the end of the time-range of a frequency rise, and we note their corresponding frequencies x_{min} and x_{max} respectively.
3. We apply the logit transformation to the frequency points between x_{min} and x_{max}:

$$y_i = log\left(\frac{x_i - x_{min}}{x_{max} - x_i}\right). \tag{4}$$

4. If the data actually follows a sigmoid function $\tilde{x}_i$ of the form:

$$\tilde{x}_i = x_{min} + \frac{x_{max} - x_{min}}{1 + e^{-hi-b}}, \tag{5}$$

then the logit transformation of this sigmoid function fits a linear function of the form:

$$\tilde{y}_i = hi + b, \tag{6}$$

which gives us the slope h, the intercept b and the residual r^2 quantifying the linear quality of the fit. Figure 6 provides an illustration of an extracted S-curve pattern of frequency rise and its corresponding logit transform.

4.3 Experimentation

Experimental Settings. We investigated the presence of S-curve patterns on a handful of words. We used the same 100 words that we have considered for the case studies of the first part of our study (See Table 2 for examples). We then used the mathematical procedure described in the previous subsection to look for S-curve patterns in the frequency profile of these words over time. We have decided to keep the investigation dataset small in order to facilitate the subsequent qualitative analysis.

Results. In our study, we wanted to select only S-curves of highly satisfying quality. We set the residual r^2 quantifying the quality of the logit fit, as explained in Subsect. 4.2, to a high value (98%) in order to ensure the S-shape quality of the extracted curves. Moreover, we restricted the extraction to S-curves that cover at least one decade in length. Still, our method enabled us to find at least one S-curve (and up to four in some cases) in the frequency evolution for 46% of the studied. A lower threshold for r^2 would have yielded many more. The word *Array*, as we have seen in Table 2 above, is a good example of semantic change, having shifted from the idea of order of battle (Webster, 1828, s.v. *array*, first entry: "Order; disposition in regular lines; as an army in battle array. Hence a posture of defense") to that of imposing numbers (Merriam-Webster, online edition, s.v. *array*, first entry: "an imposing group: large number") [34]. Figure 5 illustrates the frequency profile of this word in the past two centuries, while Fig. 6 shows the S-curve pattern of frequency rise period extracted from its profile.

Theoretically speaking, semantic change can occur in two cases: because a word gains new meanings (semantic expansion), or because it loses some (semantic reduction). Since only the former case is associated with a frequency rise, we decided to focus the extraction on words having a growing frequency trend over time (instances of semantic expansion). This brings up the percentage of S-curve pattern presence to 75%. These results suggest that the S-curve does indeed seem to be, to some extent, a generic pattern of lexical semantic change, and especially of semantic expansion. However, when questioning the universality of the S-curve pattern, the results show that it is actually not so pervasive as to be qualified as universal, at least on the basis of our investigation dataset based on frequencies extracted from the Google Books N-gram corpus. A more systematic investigation into bigger datasets, along with individual case studies, is in order if we want to get a full picture.
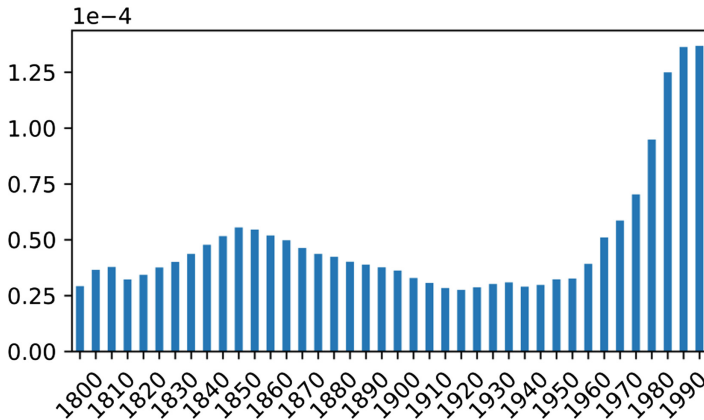
Fig. 5. Overall evolution of the relative smoothed frequency of use of the form *array* in the corpus.
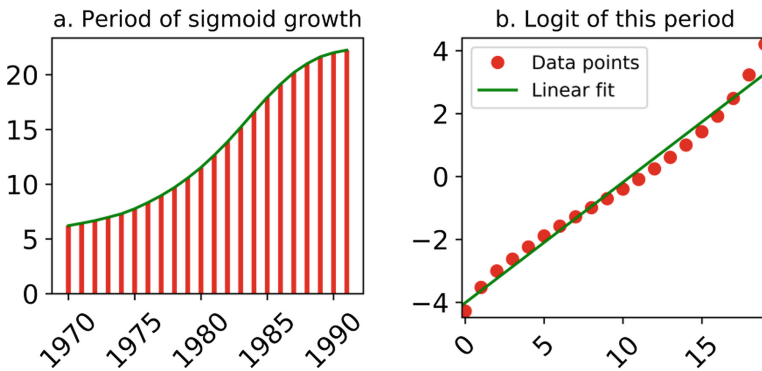
Fig. 6. Extracted S-curve pattern of frequency rise period for the form *array* (left) and its corresponding logit transform (right).

5   Conclusions

In conclusion, we presented a computational corpus-based study to analyse the semantic evolution of words over time. First, we tried to capture the systemic change of word meanings in an empirical model that is also predictive, making it falsifiable. In order to illustrate the significance of this kind of empirical model, we conducted an experimental evaluation using the Google Books N-Gram corpus. The results show that the model is partly successful in capturing the semantic change and can achieve a high degree of accuracy on predicting words' distributional semantics.

We then proposed to investigate the pervasive nature of the S-curve pattern in the case of semantic change. To do so, we used an automated procedure both to empirically extract the words that have known biggest semantic shifts in the past two centuries, and to extract S-curve patterns from their frequency profiles. Performing a statistical observation over 100 cases of semantic change from the Google Books N-gram corpus,

we established the generic character of the S-curve, especially in the case of frequency rises related to semantic expansion.

Although our experiments described in this study are still in their preliminary stages, we believe that this approach can provide linguists with a refreshing perspective on language evolution by making it possible to observe large-scale evolution in general and semantic change in particular. It thus nicely complements existing methods and reinforces a falsifiability approach to linguistics. Based on the current results, we have identified several future research directions. The RNN model that we propose to use in the first part of our study is rather standard and simplistic compared to the complexity of semantic change. We therefore intend to explore deeper networks and to put more time and effort in the fine tuning process of its hyper-parameters. On the other hand, a good question would be to analyse the possible interactions between different linguistic properties of the word such as its actual degree of semantic change and the degree of its polysemy in the time period covered by the S-curve pattern. This is only a broad perspective of research, which we shall explore more thoroughly in future works.

References

1. Simpson, J.A., Weiner, E.S.C.: The Oxford English Dictionary. Oxford University Press, Oxford (1989)
2. Dubossarsky, H., Tsvetkov, Y., Dyer, C., Grossman, E.: A bottom up approach to category mapping and meaning change. In: NetWordS, pp. 66–70 (2015)
3. Traugott, E.C., Dasher, R.B.: Regularity in Semantic Change. Cambridge University Press, Cambridge (2001)
4. Bailey, C.-J.N.: Variation and linguistic theory (1973)
5. Kroch, A.S.: Reflexes of grammar in patterns of language change. Lang. Var. Change. **1**, 199–244 (1989)
6. Steels, L.: Modeling the cultural evolution of language. Phys. Life Rev. **8**, 339–356 (2011)
7. Boukhaled, M., Fagard, B., Poibeau, T.: Modelling the semantic change dynamics using diachronic word embedding. In: Proceedings of the 11th International Conference on Agents and Artificial Intelligenc, ICAART 2019. Prague, Czech Republic (2019)
8. Kim, Y., Chiu, Y.-I., Hanaki, K., Hegde, D., Petrov, S.: Temporal analysis of language through neural language models. arXiv Preprint. arXiv:1405.3515 (2014)
9. Rosin, G.D., Radinsky, K., Adar, E.: Learning Word Relatedness over Time. arXiv Preprint arXiv:1707.08081 (2017)
10. Szymanski, T.: Temporal word analogies: identifying lexical replacement with diachronic word embeddings. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 448–453 (2017)
11. Kutuzov, A., Velldal, E., Øvrelid, L.: Temporal dynamics of semantic relations in word embeddings: an application to predicting armed conflict participants. arXiv Preprint arXiv:1707.08660 (2017)
12. Hamilton, W.L., Leskovec, J., Jurafsky, D.: Diachronic word embeddings reveal statistical laws of semantic change. arXiv Preprint arXiv:1605.09096 (2016)

13. Feltgen, Q., Fagard, B., Nadal, J.-P.: Frequency patterns of semantic change: corpus-based evidence of a near-critical dynamics in language change. R. Soc. Open Sci. **4**, 170830 (2017)
14. Turney, P.D., Pantel, P.: From frequency to meaning: vector space models of semantics. J. Artif. Intell. Res. **37**, 141–188 (2010)
15. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci. **41**, 391–407 (1990)
16. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
17. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv Preprint arXiv:1301.3781 (2013)
18. Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
19. Lin, Y., Michel, J.-B., Aiden, E.L., Orwant, J., Brockman, W., Petrov, S.: Syntactic annotations for the Google books ngram corpus. In: Proceedings of the ACL 2012 System Demonstrations, pp. 169–174 (2012)
20. Hamilton, W.L., Leskovec, J., Jurafsky, D.: Cultural shift or linguistic drift? comparing two computational measures of semantic change. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, p. 2116 (2016)
21. Bengio, Y.: Markovian models for sequential data. Neural Comput. Surv. **2**, 129–162 (1999)
22. Lafferty, J., McCallum, A., Pereira, F.C.N.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data (2001)
23. Medsker, L.R., Jain, L.C.: Recurrent neural networks. Des. Appl. **5** (2001)
24. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**, 1735–1780 (1997)
25. Pascanu, R., Mikolov, T., Bengio, Y.: On the difficulty of training recurrent neural networks. In: International Conference on Machine Learning, pp. 1310–1318 (2013)
26. Iacobacci, I., Pilehvar, M.T., Navigli, R.: Sensembed: learning sense embeddings for word and relational similarity. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 95–105 (2015)
27. Rogers, E.M.: Diffusion of Innovations. Simon and Schuster, New York (2010)
28. Denison, D.: Logistic and simplistic S-curves. Motiv. Lang. Chang. **54**, 70 (2003)
29. Labov, W.: Principles of Linguistic Change, Volume 3: Cognitive and Cultural Factors. Wiley, Oxford (1994)
30. Ghanbarnejad, F., Gerlach, M., Miotto, J.M., Altmann, E.G.: Extracting information from S-curves of language change. J. R. Soc. Interface **11**, 20141044 (2014)
31. Nevalainen, T.: Descriptive adequacy of the S-curve model in diachronic studies of language change. In: Can We Predict Linguistic Change? (2015)
32. Blythe, R.A., Croft, W.: S-curves and the mechanisms of propagation in language change. Language (Baltim) **88**, 269–304 (2012)
33. Feltgen, Q.: Statistical physics of language evolution: the grammaticalization phenomenon (2017)
34. Webster, N.: Noah Webster's first edition of an American dictionary of the English language. Foundation for Amer Christian (1828)