Letters from the past: modeling historical sound change through diachronic character embeddings

Anonymous ACL submission

Abstract

While a great deal of work has been done on NLP approaches to Lexical Semantic Change detection, other aspects of language change have received less attention from the NLP community. In this paper, we address the detection of sound change through historical spelling. We propose that a sound change, $a \rightarrow b / c$, can be captured by comparing the relative distance through time between the distributions of the corresponding characters, *a* and *b*. We model these distributions using PPMI character embeddings. We verify this hypothesis in synthetic data and then test the method's ability to trace the well-known historical change of lenition of plosives in Danish historical sources. We show that the models are able to identify several of the changes under consideration and to uncover meaningful contexts in which they appeared. The methodology has the potential to contribute to the study of open questions such as the relative chronology of sound shifts and their geographical distribution.

1 Introduction

The study of sound change goes back to the beginnings of modern linguistics in early nineteenth century, when scholars such as Rask and Grimm started making hypotheses about the way sound changes over time, which in turn lead to the discovery of regular sound correspondences between ancient languages and the identification of cognates in modern ones (Murray, 2015).

Since spoken language from the past is not available, however, sound change in ancient languages must be deduced from written records by considering development in spelling through time. In addition, while we may be able to see from the written records that a change did occur, less is known on the specific dynamics of the change. Details of these dynamics include knowledge of when the change started to appear, how long it took for it to be complete, what was the relative chronology of individual sounds in a larger shift, what was the geographical distribution of a change and so forth.

Due to the sparsity of linguistic evidence, detailed empirical studies of chronological sound change are difficult to conduct. This is especially the case for older stages of languages, where little written text was produced, and much of what did exist has been lost in transmission. However, as we move forward in history to the rise of bureaucracy, for example in medieval Europe, we see that an extensive amount of written records were made. Text from this period of time is interesting in the context of a study of sound change because it shows great variability in spelling patterns. With the digitalization of such archives[1], therefore, new opportunities arise to apply computational methods to the study of sound change through written text.

Considerable effort has already been devoted to the development of computational approaches aimed at discovering lexical semantic change (LSC) in historical corpora. However, change related to phonology, morphology and syntax has remained out of the spotlight in NLP research. In this study, we seek to bridge this gap as regards phonology: Inspired by the work on LSC, we propose a method whereby sound change is traced via spelling change in historical text and modeled by training diachronic character embeddings over text from different time periods.

We start by reviewing previous approaches to the automatic detection of semantic shifts and spelling modification due to sound change. Then we formulate our hypothesis that a sound change can be traced using diachronic distributional embeddings. While sound change is not completely analogous to word meaning change, we argue that similar methods can be used for both. To verify our hypothe-

[1] A list of available resources for different languages is provided in the Guide to Medieval Manuscript Research from the University of Chicago Library: `https://guides.lib.uchicago.edu/c.php?g=813534&p=5805534`.

sis, we conduct three studies on simulated sound change. First, we test the methods on the phonological environment of a simple artificial language. Then, we apply the same methods to a more complex scenario created by simulating sound change in a corpus of synchronic Danish text. Having established the suitability of the methods on these two datasets, we finally experiment with tracing a well-known sound change in real historical language data, again in Danish.

2 Related work

The application of NLP methods to automatic LSC detection is already a rather well-developed subfield of NLP research (Tahmasebi et al., 2018; Kutuzov et al., 2018). In particular, the emergence of word embeddings as a viable way to model the distributional hypothesis in semantics (Firth, 1957) has paved the way for an application of word embeddings to LSC modeling (Kim et al., 2014; Hamilton et al., 2016b; Eger and Mehler, 2016; Yao et al., 2018). Synchronically, the meaning of a word is characterized by word embeddings in terms of the contexts it appears in. LSC is captured by training word embeddings at different time points and comparing these distributions typically using cosine distance.

The main issues in this comparison is the alignment of temporal embeddings spaces, especially for neural embeddings as these are initialized and trained stochastically, which means that separate runs – on even the same data – will yield different embeddings spaces. Thus, work has focused on the development of methods to perform alignments to make embedding spaces comparable across time (see Kutuzov et al. (2018) for an overview). As an alternative to neural embeddings, scholars have also used purely count-based measures, which are naturally aligned across dimensions. Normalisation techniques are also applied, e.g. based on positive pointwise mutual information (PPMI) (Hamilton et al., 2016b; Yao et al., 2018).

Most studies of LSC do not rely on a control dataset against which to validate their conclusions. In Dubossarsky et al. (2017), on the contrary, it is argued that any claims about putative laws of semantic change in diachronic corpora must be evaluated against a relevant control condition. The authors propose a methodology in which a control condition is created artificially from the original diachronic text collection by reshuffling the data. No systematic LSC is expected in the artificially developed control dataset.

The distributional hypothesis has also been proposed as an explanatory model within the domain of phonology suggesting that phonological classes are acquired through distributional information (Chomsky and Halle, 1968; Mielke, 2008). Driven by this hypothesis, recent work has focused on testing how distributional properties can be learned by phoneme embeddings (see Mayer 2020 for an overview). Silfverberg et al. (2018) investigated to what extent learned vector representations of phonemes align with their respective representations in a feature space in which dimensions are articulatory descriptors (e.g., ±plosive). Recently, Mayer (2020) has shown that phonological classes, such as long and short vowels, can be deduced from phoneme embeddings normalised using PPMI by iteratively performing PCA on candidate classes.

Thus, while the distributional hypothesis for phonology is well-established, one notable issue is the fact that the empirical evidence to study sound change is relatively inaccessible since it requires recorded speech or phonologically transcribed data. Simulation is therefore used as a tool for studying the underlying mechanisms of sound change by creating computational models based on linguistic theory (Wedel, 2015). Through simulation, questions pertaining to e.g., what factors influence the (in)stability of vowel systems across generations (de Boer, 2003) can be modeled by controlling the assumptions made by the model. Work on simulation ranges from implementing theoretical approaches using mathematical models (Pierrehumbert, 2001; Blythe and Croft, 2012) to iterated learning and neural networks (Hare and Elman, 1995; Beguš, 2021).

While the output of such models can be tested empirically on what we observe at a synchronic level, they are primarily theoretically driven. In this paper, we wish to take a data-driven approach and utilize some of the methods reviewed above to track historical sound change in writing. Rather than using word embeddings as done to model lexical change, we will use character embeddings, that are better suited to the task of sound change modeling.

3 Modeling sound change

Within the field of LSC detection, change in word semantics is traditionally measured by computing *pairwise similarity* (Hamilton et al., 2016b) over a time series, (t, ..., $t+\delta$), in which a shift in the meaning of a word, w_i, can be measured by its relative distance to another word, w_j. In this way, hypotheses about specific shifts may be tested. Another measure is *semantic displacement*, in which semantic change for a given word is quantified by measuring its temporal displacement. For both measures, looking at consecutive time steps provides a measure to the rate of change of a word – in relation to another word, or independently.

While LSC is about meaning shifts of unchanged word forms, sound change is a change of form, i.e., a given phoneme changes to another one within certain contexts. We denote such a change $a \rightarrow b$ / c. While changes of either a or b will be reflected in changes to their individual distributions (*displacement*), looking at them independently of one another will not tell us whether one of the phonemes is becoming similar to the other. Therefore, we suggest to look at the *pairwise similarity* between a and b. More specifically, given a time series (t_1, ..., t_n), in which t_1 denotes a time before a sound change was in effect and t_n denotes a time where a sound change is completed, we expect b_i to *move* towards a_1 as $i \rightarrow n$, in other words to become similar to a_1, since it will begin to appear in the same contexts.

As was noted earlier, sound is not accessible in historical text, to which we would like to be able to apply our methodology. Therefore, we take graphemes as a proxy for sound, and model sound change through changes in the distance between pairs of character distributions. In addition, before assuming that an observed decrease in the distance between two such distributions reflects a real change, we also want to see that the same decrease is not visible in a control corpus in which no such change has indeed taken place.

4 Experimental setup

In order to verify the hypothesis that sound change can be traced using distributional information with the methodology proposed above, we test whether we are able to trace simulated change in synthetic data. As a first synthetic setting, we restrict ourselves to track change in a synthetic language with simple phonotactics. In this way, we get a sense of whether the proposed hypothesis works under perfect conditions, i.e., one in which characters correspond with phonemes one-to-one . In the second synthetic setting, we seek to imitate the condition of tracing change in an orthographic setting by simulating change in a corpus of synchronic text in which character distributions interact with the noise added by spelling and lexicon.

In both synthetic settings, we compare the simulated change to a control setting where no change has occurred.

Finally, we will test the hypothesis on real data. Our goal is to trace the lenition after vowels of voiceless plosives, *p t k*, to their voiced counterparts, *b d g*, in historical Danish. While this change is believed to have initiated around the beginning of the 14$^{\text{th}}$ century, details about the relative chronology of the series and geographical distribution of the change are difficult to account for (Frederiksen, 2018). Therefore, in an attempt to discover interesting patterns of this change, we train character embeddings on historical sources from the periods following the time when the change is believed to have started. As we did for the synthetic data, and again following Dubossarsky et al. (2017), we also introduce a control setting to test the significance of the observed changes.

4.1 Data

Parupa is an artificial language introduced by Mayer (2020). It is characterized by a small phonological inventory[2], and simple phonotactic rules for how sounds combine:

- only *CV* syllables are allowed
- */p t k/* occur before high vowels, */i u/*
- */b d g/* occur before non-high vowels, */e o/*
- only */b p/* occur word-initially
- */r/* occurs before all vowels
- all consonants can occur before */a/*

We created five corpora of Parupa using the Hidden Markov Model provided by Mayer (2020)[3]: While the first corpus, $\texttt{parupa}_1$, preserves the phonotactic rules listed above, the remaining four include a sound change, $p \rightarrow b$ /_ u, i [4] which happens gradually (linearly) and is fully completed in $\texttt{parupa}_5$. Additionally, we created five control corpora (one for each of the target ones and with the same vocabulary) which do not include any

[2] *C*: */p t k b d g r/ V*: */i e u o a/*

[3] https://github.com/connormayer/distributional_learning

[4] i.e., *p* changes into *b* when preceding *u* or *i*.

simulated sound change. Each of the corpora consists of $50,000$ words.

The Danish UD treebank To collect a corpus of synchronic language, we extracted the training sentences from the Danish UD treebank (Johannsen et al., 2015). From this collection of sentences, we created five sub-corpora (`UD-Danish`$_{1\text{-}5}$) in which we simulated a sound change, *g* $\rightarrow$ *k / V_{V # t#}*[5]. As done in the case of Parupa, the sound change was simulated gradually, with linear increase in change probabilities (i.e., $0, 0.25, 0.50, 0.75, 1$). To create the control condition, we also kept a version of the sub-corpora where no change was simulated. The five control versions are thus identical to the five target corpora in terms of vocabulary and distributions, except for the simulated change.

Historical spellings of geographical names *Danmarks Stednavne* is a on-going lexicographic book series creating a register of geographical names in Denmark. The register also serves as a philological resource by listing attestations of the names coming from various historical resources. For example, the entry for *Copenhagen* includes over 700 historical attestations listed by date[6]. In addition to the printed volumes (Danmarks Stednavne, 1922–2013), geographical names and their connected metadata (e.g., geographical location and historical attestations) have been digitized, and can be found in an online edition[7] which comprises over $210,000$ names and $900,000$ historical attestations. To study the lenition of */p t k/*, we extracted historical attestations of names ranging from the 12th to the 18th century. Using the attestation before the 14th century as a reference to the time before the change was initiated (t_1), we divided the list of names into bins of half a century to track the development of character embeddings through time[8], This provides us with eleven sub-corpora with $31,000$ ($\pm 15,000$) name tokens on average.

[5] i.e., *g* between vowels, word-final after vowel, or after vowel preceding word-final *t*. The latter condition was created in order to capture adverbial forms of adjectives ending in *-g*.

[6] e.g., *Kopmanahafn* (1247), *Køpmannehafn* (1249), *Kiøpnehaffn* (1388), *Kiøbendehaffn* (1429).

[7] `https://danmarksstednavne.navneforskning.ku.dk`

[8] The choice of bin size is an important issue (Kutuzov et al., 2018). From a philological perspective, 50 years correspond to two generations of writers ('spellers'), which is considered a realistic bin size to track development in writing.

In order to create a control setting, we generated a corresponding number of sub-corpora by stratifying the names with respect to their date of attestation, following the approach by Dubossarsky et al. (2017).

4.2 Character embedding model

To represent characters in a distributional space, we use PPMI embeddings. Contrary to dense embeddings, these are easy to interpret and when compared across different initializations, they are naturally aligned, so we do not introduce noise caused by the alignment process.

For the synthetic settings, we limit the context windows to the modeling of trigrams, which should be sufficient to model the context of where a sound change occurs. For the tracking of lenition in Danish, we expand to context to 4-grams. Using the implementation by Mayer (2020), the sliding window is directional, and thus we distinguish contexts preceding and following the target character. While this directionality is neglected when creating PPMI word embeddings, the direction matters when using character embeddings to test the intuition behind the distributional hypothesis, in which direction in a context is meaningful.

4.3 Measuring change

We measure sound change in terms of a decrease in the distance between two character distributions over time. In other words, given two character distributions A and B corresponding to any two phonemes */a/* and */b/*, we should see that $distance(A^{(1)}, B^{(n)})$ gets smaller for greater values of *n* if there is a change $A \rightarrow B$.

While most studies use cosine distance to measure the difference between distribution (Kutuzov et al., 2018), we chose to use Euclidean distance as it directly reflects our hypothesis by taking the sum of differences in each dimension (context).

For each of the corpora being investigated, we use the R software (R Core Team, 2021) and the 'effects' package (Fox and Weisberg, 2019) to build linear regression models that predict the distributional distance between two sounds per temporal interval in the target and the control versions of the corpus. The advantage of employing linear regression in this case is that we can test the effect of multiple factors as well as their interaction[9]. In

[9] See also (Shoemark et al., 2019) on the advantages of using linear regression in semantic change detection.

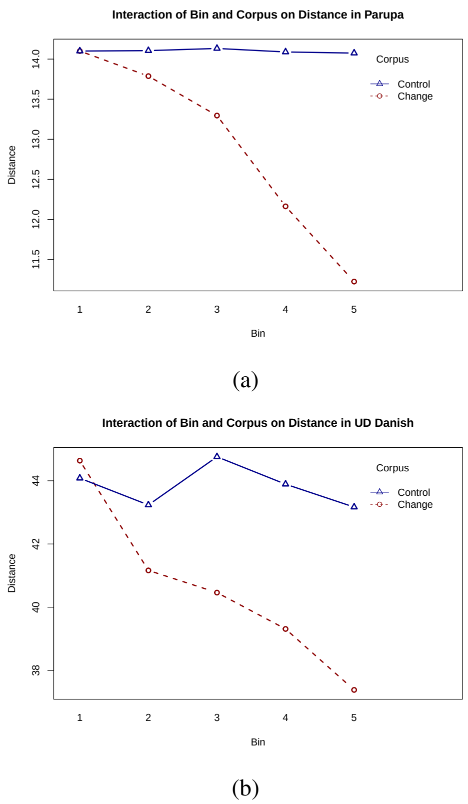

Figure 1: Interaction of Bin and Corpus on Distance in Parupa (a) and the Danish UD treebank (b)

our case, distance between the two sounds being investigated is the dependent variable, and we want to predict main effects of temporal interval and corpus as well as the interaction between them. To argue that there has been a sound change across time, there must be a significant effect of temporal interval on distance. In addition, we would like to see an interaction between this effect and the effect of the corpus variable in that the change should be absent, or at least significantly smaller, in the control corpus.

5 Results

Table 1 shows the results of the linear regression models we developed to test whether any evidence of sound change discovered in the target corpora, where sound change is either simulated or historically present, stands the comparison with the control corpora. The 'intercept' estimate corresponds to the distance predicted between the two sounds being investigated in the initial temporal interval. The 'Bin' effect shows by how much the distance is expected to change for every temporal interval. A negative effect means that the distance between the two sounds is becoming smaller. The 'Control' ef-

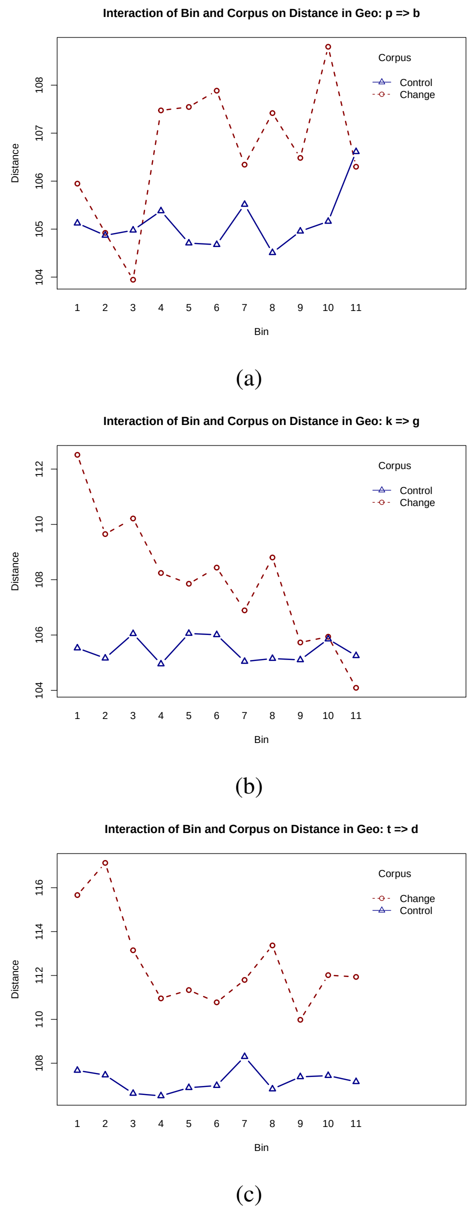

Figure 2: Interaction of Bin and Corpus on Distance in the Danish Geographical Names: Looking at p $\rightarrow$ b (a), k $\rightarrow$ g (b) and t $\rightarrow$ d (c)

	Effect	Estimate	Std. Error	p-value
Parupa	(Intercept)	15.13	0.23	<.001 ***
	Bin	-0.74	0.07	<.001 ***
	Control	-1.01	0.33	<.05 *
	Bin:Control	0.73	0.10	<.001 ***
UD Danish	(Intercept)	45.50	0.80	<.001 ***
	Bin	-1.64	0.24	<.001 ***
	Control	-1.32	1.13	0.28
	Bin:Control	1.52	0.34	<.01 **
Geo Names p → b	(Intercept)	105.36	0.63	<.001 ***
	Bin	0.21	0.09	<.05 *
	Control	-0.64	0.89	0.48
	Bin:Control	-0.14	0.13	0.28
Geo Names k → g	(Intercept)	111.86	0.52	<.001 ***
	Bin	-0.64	0.08	<.001 ***
	Control	-6.28	0.74	<.001 ***
	Bin:Control	0.62	0.11	<.001 ***
Geo Names t → d	(Intercept)	114.92	0.87	<.001 ***
	Bin	-0.39	0.13	<.01 **
	Control	-7.81	1.23	<.001 ***
	Bin:Control	0.41	0.18	<.05 *

Table 1: Coefficients of linear regression models predicting increase of distance between the investigated sounds in two simulated corpora.

fect shows the predicted change to the initial Intercept in the control corpus, and finally ‘Bin:Control’ shows the interaction between temporal bin and corpus type.

In both the corpora where change is simulated, there is a significant effect of temporal interval. This is expected given the fact that gradual change has been induced in the data. The effect of the control corpus on the initial sound distance is significant for Parupa but not for UD Danish. More importantly, the interaction between the effect of the temporal bin and the control corpus is significant in both cases. The interaction supports the hypothesis that we see a pattern of change in the simulated corpora that is significantly different compared to the control data. The interactions are shown in the plots in Figure 1.

Turning to the results for the Danish Geographical Names corpus, while the models show significant effects of Bin, Control and interaction between the two for the $k \rightarrow g$ and the $t \rightarrow d$ changes, only the effect of Bin is significant for the $p \rightarrow b$ change. When we look at the corresponding interaction plots in Figure 2, we see that the distance between p and b in the corpus seems to increase rather than diminish (as also shown by the positive Bin effect), and to do so in a rather non-linear way. The changes displayed in the plots in (b) and (c), on the contrary, follow the expected trend. The observed consonant is moving towards its voiced version in the real corpus but not in the control.

6 Discussion

The results from the two simulation studies suggest that sound change can be traced with our proposed methodology of measuring the distance between pairs of character distributions over time. We showed this both in a simplified setting (Parupa), and in the orthographically noisy environment provided by synchronic Danish data (UD Danish).

The main assumption in these simulation studies was that change could be modeled linearly. However, as discussed by scholars, change is most often not linear, but rather follows an s-shaped curve through a community (Denison, 2003). In a similar study on synthetic data, nevertheless, Shoemark et al. (2019) showed for LSC detection that tracing the change under a linear assumption, such as ours, still performs well. The results obtained in our study seem to confirm this finding in the case of sound change.

Moving on to the results on the tracing of lenition in historical sources, we were able to identify a change from $/t\ k/ \rightarrow /d\ g/$. However, this general result does not tell us much about what patterns the model picked up. To get a sense of this, instead of looking at the euclidean distance for the full embedding, we ran linear regression on the target data looking at differences between character distributions for each dimension. We then extracted the patterns corresponding to the dimensions showing significant differences and considered those with the highest Pearson's r coefficient (Tables 2-4).

Starting with the resulting patterns for Parupa and UD Danish, in both cases we are able to identify the exact contexts where the change was simulated: In Parupa before *i/u* and in the UD Danish corpus, between vowels and in the frequent suffix *-ig(t)* (although the end-of-word is not captured due to n-gram size restrictions).

Moving on to the tracing of sound change in real data, we focus our analysis on $k \rightarrow g$, which showed the greatest change. Considering first the word-final patterns, `wi_#` and `vi_#`, to spellings of the word *vig* 'inlet', commonly used as a suffix in the formation of geographical names in Danish. Descending from a Proto-Germanic word with final *-k* (*wīkwan* 'to give way; to turn (away)', compare German *weichen* 'id.' and Dutch *wijken* 'id.' (Kroonen, 2013)), the suffix is in early sources attested with a *-k*: For example, out of the six written sources of the geographical name *Rørvig* before the 14th century (corresponding to bin 1-3 in our study), four were written with a *-k*, while in later sources forms with *-g* became predominant, with the latest attestation of *-k* appearing in 1465. All of the patterns can be attributed to spellings related to similar changes[10,11] with the exception of `oli_` and `ri_#`, which do not have comparable ancestors with *-k*: These will have to be explained by later innovations, the first by the emergence of the word *bolig* 'home;dwelling' in geographical name formation, and the latter possibly indicating later spellings of names ending *-rg*.

This latter example is related to an important issue in language evolution: When language language changes through generations, we also observe shifts in culture. Different types of 'data drift' are in fact discussed by Hamilton et al. (2016a) in the context of LSC. The authors suggest that they may be modeled independently of each other by means of different measures of change. The effect of cultural change has yet to be discussed for sound change. However, it is an important discussion, since phonology, when looking at it from a corpus-based perspective, is not only governed by phonotactic constraints, but also a by-product of word usage, which is in turn dependent on cultural patterns.

In this respect, another important point to note about the retrieved patterns – both from the simulation of UD Danish and the tracing of $k \rightarrow g$ – is that many of them reflect derivational or inflectional suffixes, and are thus characterized by high frequency of occurrence across word forms. While the observation that frequent patterns are more easily captured may seem trivial, it cannot be ignored that the model may be less sensitive to infrequent patterns.[12]

The same mechanism is reflected in the model's lack of generalisation, which explains treating forms like *vig*, *wig* and *viig* as separate entries. This is a design consequence, in that we use PPMI weighting on raw n-gram counts. This method enabled us to interpret the exact inner workings of the model and find the contexts in which a change has happened. If we had used neural methods, in which characters are represented by dense embeddings, similar characters would have shared similar representations, thereby perhaps allowing the model to generalise e.g., to sound change occurring after a *vowel*. In this study, we wanted to privilege explainability, but dense representations should be explored in the future.

7 Conclusion and future work

In this paper we presented a novel method for the modeling of sound change through the use of diachronic character embeddings. Sound change is modeled in terms of increasing similarity between character distributions across time intervals. The proposed method was tested on synthetic data with promising results, and then applied to a real world scenario with the goal of tracing the lenition of

[10] Danish *sig* 'bog; mire' from Old Danish *sik*, compare Norwegian and Swedish (dialectal) *sik* (Danmarks Stednavne, 1922–2013)

[11] Danish *ager* 'field' from Proto-Germanic *akra*, compare English *acre* and Swedish *åker* (Kroonen, 2013).

[12] Whether frequency could explain the lack of evidence for observing $p \rightarrow b$ is to be investigated further. Germanic p descends from Proto-Indo-European (PIE) *b, which has a special place in the PIE phoneme inventory, being the black sheep that some scholars do not believe to have existed due to its few attestations. Thus, the attestations of Germanic p most often come from loan words and are not seen in morphemes. Thus the evidence for $p \rightarrow b$ is inherently scarcer.

4-gram	Slope	Pearson's r
i_i	-0.30	-0.93
o_u	-0.29	-0.96
_u#	-0.29	-0.95
a_u	-0.28	-0.95
e_i	-0.28	-0.94

Table 2: Analysis of the simulated change from p to b in Parupa. Five most important dimensions after filtering 3-grams with respect to Pearson's r (<-0.2) and p-value(<0.05). The table is ordered by slope. '#' indicates word boundaries.

4-gram	Slope	Pearson's r
li_	-0.71	-0.92
i_e	-0.63	-0.89
i_t	-0.58	-0.93
di_	-0.58	-0.97
a_e	-0.57	-0.95

Table 3: Analysis of the simulated change from g to k in synchronic Danish. Five most important dimensions after filtering 3-grams with respect to Pearson's r (<-0.2) and p-value(<0.05). The table is ordered by slope. '#' indicates word boundaries.

/p t k/ → */b d g/* in Danish by looking at spelling in historical sources. The method was able to detect the changes for two of the sound pairs, and also to point at specific contexts of occurrence that influenced the changes.

For scholars interested in sound change, there are a number of important open questions, such as the relative chronology and geographical distribution of sound shifts. Although we have not addressed these questions here, we believe our methodology can be further developed in ways that would allow to do so, e.g., by adding geographical location as an additional factor in the models. Both issues would constitute interesting avenues for future research.

In this paper we have used purely count-based methods. While this approach enables us to directly interpret the results of the models, it also suffers from its inability to generalise across contexts. This drawback motivates experimenting with neural methods that make use of dense character representations, to test whether they can make similar generalisations as done by historical linguists, particularly as regards infrequent patterns that could be captured across word forms.

4-gram	Slope	Pearson's r
rvi_	-0.50	-0.85
sii_	-0.45	-0.79
æ_er	-0.43	-0.78
m#a_	-0.40	-0.80
vi_#	-0.40	-0.79
oli_	-0.40	-0.84
ri_#	-0.38	-0.83
wi_#	-0.31	-0.83
ara_	-0.31	-0.62
a_re	-0.28	-0.71

Table 4: Analysis of the change from k to g in historical records of geographical names. Ten most important dimensions after filtering 4-grams with respect to Pearson's r (<-0.2) and p-value(<0.05). The table is ordered by slope. '#' indicates word boundaries.

References

Gašper Beguš. 2021. Deep sound change: Deep and iterative learning, convolutional neural networks, and language change.

Richard A Blythe and William Croft. 2012. S-curves and the mechanisms of propagation in language change. *Language*, pages 269–304.

Noam Chomsky and Morris Halle. 1968. *The sound pattern of English*. Harper & Row, New York.

Danmarks Stednavne. 1922–2013. *Danmarks Stednavne*. C. A. Reitzel, København.

Bart de Boer. 2003. Conditions for stable vowel systems in a population. In *Advances in Artificial Life*, pages 415–424, Berlin, Heidelberg. Springer Berlin Heidelberg.

David Denison. 2003. Log(ist)ic and simplistic s-curves. In Raymond Hickey, editor, *Motives for Language Change*, page 54–70. Cambridge University Press.

Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman. 2017. Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1136–1145, Copenhagen, Denmark. Association for Computational Linguistics.

Steffen Eger and Alexander Mehler. 2016. On the linearity of semantic change: Investigating meaning variation via dynamic graph models. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 52–58, Berlin, Germany. Association for Computational Linguistics.

John R Firth. 1957. A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*.

John Fox and Sanford Weisberg. 2019. *An R Companion to Applied Regression*, third edition. Sage, Thousand Oaks CA.

Britta Olrik Frederiksen. 2018. Gammeldansk. In Ebba Hjorth, Henrik Galberg Jakobsen, Bent Jørgensen, Brigitte Jacobsen, Merete Korvenius Jørgensen, and Laurids Kristian Fahl, editors, *Ord for ord for ord*, volume 2 of *Dansk Sproghistorie*, pages 151–179. Aarhus Universitetsforlag.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016a. Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2116–2121, Austin, Texas. Association for Computational Linguistics.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016b. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Mary Hare and Jeffrey L Elman. 1995. Learning and morphological change. *Cognition*, 56(1):61–98.

Anders Johannsen, Héctor Martínez Alonso, and Barbara Plank. 2015. Universal dependencies for danish. In *International Workshop on Treebanks and Linguistic Theories (TLT14)*, page 157.

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal analysis of language through neural language models. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 61–65, Baltimore, MD, USA. Association for Computational Linguistics.

Guus Kroonen. 2013. *Etymological Dictionary of Proto-Germanic*. Brill, Leiden, The Netherlands.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: a survey. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1384–1397, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Connor Mayer. 2020. An algorithm for learning phonological classes from distributional similarity. *Phonology*, 37(1):91–131.

Jeff Mielke. 2008. *The emergence of distinctive features*. Oxford University Press.

Robert W Murray. 2015. The early history of historical phonology. In Patrick Honeybone and Joseph Salmons, editors, *The Oxford Handbook of Historical Phonology*, Oxford Handbooks in Linguistics. OUP Oxford.

Janet B. Pierrehumbert. 2001. *Exemplar dynamics: Word frequency, lenition and contrast.*, Typological studies in language, vol. 45., pages 137–157. John Benjamins Publishing Company, Amsterdam, Netherlands.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Philippa Shoemark, Farhana Ferdousi Liza, Dong Nguyen, Scott Hale, and Barbara McGillivray. 2019. Room to Glo: A systematic comparison of semantic change detection approaches with word embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 66–76, Hong Kong, China. Association for Computational Linguistics.

Miikka P. Silfverberg, Lingshuang Mao, and Mans Hulden. 2018. Sound analogies with phoneme embeddings. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2018*, pages 136–144.

Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2018. Survey of computational approaches to lexical semantic change. In *Preprint at ArXiv 2018.*

Andrew Wedel. 2015. Simulation as an investigative tool in historical phonology. In Patrick Honeybone and Joseph Salmons, editors, *The Oxford Handbook of Historical Phonology*, Oxford Handbooks in Linguistics. OUP Oxford.

Zijun Yao, Yifan Sun, Weicong Ding, Nikhil Rao, and Hui Xiong. 2018. Dynamic word embeddings for evolving semantic discovery. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*.