“Vaderland”, “Volk” and “Natie”: Semantic Change Related to Nationalism in Dutch Literature Between 1700 and 1880 Captured with Dynamic Bernoulli Word Embeddings

Marije Timmermans
STUDENT NUMBER: 2042487

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
DEPARTMENT OF COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

Thesis committee:

Dr Eva Vanmassenhove
Dr Yash Satsangi

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
January 2021

Preface

I would like to thank Imme Wagemans, Giandrikc Dabian and Erika Pauliukevičiūtė for the weekly meetings that made me feel like I wasn't alone in this during these weird Covid times. Special thanks to dr Eva Vanmassenhove for guidance and providing the idea for my thesis. A huge thanks to Luke Lefebure, for providing an easier application of the Dynamic Bernoulli Embedding model.[1] And thanks to Nicole Yedica for proof-reading. Finally, I would like to thank my boyfriend Paul for encouraging me to enroll in this master's program and supporting me mentally and technically during the writing of my thesis. In addition I would like to thank my parents for supporting me all these years.

1 https://github.com/llefebure/dynamic_bernoulli_embeddings

"Vaderland", "Volk" and "Natie": Semantic Change Related to Nationalism in Dutch Literature Between 1700 and 1880 Captured with Dynamic Bernoulli Word Embeddings

Marije Timmermans

In this thesis the development of nationalism will be researched by analyzing how the words "natie" (nation), "volk" (people) and "vaderland" (fatherland) have changed in Dutch fiction literature from 1700 to 1880. Earlier research has proposed diachronic or dynamic word embedding models for this purpose. The most important finding of this thesis is that Dynamic Bernoulli Word Embeddings can show and quantify a semantic shift in Dutch fiction literature. The nearest neighbors of the target words suggest that the context in which the target words are used becomes more nationalistic over time.

1. Introduction

Historical studies have shown that a nation, with one people with the same national identity, is created and is not an unchangeable element in society. The creation of a nation is done by nation building (Anderson 1983).

An example of nation building is the reforms during the short period of the Dutch Batavian Republic (1795-1801). In 1795, six years after the French Revolution, a revolutionary group, the patriots, supported by the French, changed the name of the Dutch Republic to the Batavian Republic. The name refers to the Germanic Tribe Batavi, a Germanic tribe on present day Dutch soil that is described by the Romans. The Batavian Republic strengthened the feeling of Dutch national identity by celebrating national holidays, reforming public education with an emphasis on Dutch history and Dutch language, and reinforcing the national origin myth of the Batavi as the ancestors of the modern Dutch citizens (Kloek and Mijnhardt 2001; Van Sas 1999a).

This short part of Dutch history is just one example of many nation building attempts. The first half of the nineteenth century is characterized by political events, such as the incorporation to the French Empire, regaining independence, the fusion of the former Southern and Northern Netherlands and the Belgian Revolution of 1830, resulting in a constant need for re-branding of the national identity (Van den Berg 1999).

Already before these political events, the Dutch enlightenment shaped the ideas of a Dutch national identity. Around the middle of the eighteenth century, books with "vaderland" ("fatherland") in the title were frequently published and often dealt with the typical Dutch characters (Kloek 1999).

This period in Dutch history is characterized by what Kosseleck in Brunner et al. (1972) calls the Sattelzeit. While the French Revolution of 1789 is often marked as the starting point of the modern era, the development of the modernist mindset did not happen overnight. The Sattelzeit embodies the transition period of the early modern period to the modern period, from roughly 1750 to 1850, in which European society went through a rapid cultural change. During this period, the reading public expanded, people became used to thinking historically and thinking about the future, ideologies such as nationalism arose, and abstract concepts became more politically applicable.

The aim of this thesis is to study the development of nationalism during the Sattelzeit in Dutch society by researching fiction literature from 1700 to 1880. The research question is that will be answered is: “How did the concepts “vaderland”, “volk” and “natie” undergo a semantic shift in Dutch fiction literature between 1700 and 1880?”

Studying the cultural thought of nationalism is being conducted with a Dynamic Bernoulli Word Embedding model. A dynamic word embedding model is an upcoming method in digital humanities that aims to measure how concepts have changed over time (Wevers and Koolen 2020). In this thesis, word embeddings are used to assess how the context of three target words related to nationalism have changed from the beginning of the eighteenth century to the end of the nineteenth century. These target words are “natie” (“nation), “volk” (“people”) and “vaderland” (“fatherland”).

The idea is that these word embeddings can show in which context the target words are often used, and thus can give an idea of perceptual change of these contexts. One way to see how contexts have changed is by looking at nearest neighbors of the target words, which are words that appear in a similar context as the target words do. These words are positionally close to the target word in the word embedding. In a dynamic word embedding, these nearest neighbors change over time and can tell us something about how the contexts of the target words shift over the decades.

A second, more quantitative way of looking at how the target words have changed over time is looking at the absolute drift of the target word’s embedding. The absolute drift is measured by taking the euclidean distance between the word’s embedding of the first time slice and the last time slice.

Summarizing, this research combines the historical debate on the origins of nationalism with dynamic word embeddings. By analyzing the context in which the target words related to nationalism have been used in literature and measuring the change in position in the target word’s embedding, we aim to establish whether there is a change that coincides with the upcoming cultural and political thoughts of the eighteenth and nineteenth century.

The most important finding of this thesis is that Dynamic Bernoulli Word Embeddings can show and quantify the semantic shifts of the words "vaderland", "volk" and "natie" in Dutch fiction literature. The word "natie" did not show a large semantic drift, while the words "vaderland" and "volk" show a slightly larger semantic shift. The nearest neighbors of the target words suggest that this semantic shift is indeed in the direction of a more nationalist context.

2. Related Work

Because of the multidisciplinary nature of this thesis, first the historical field of conceptual history will be described in section 2.1. In section 2.2, an overview will be given of the word embeddings, how they are used in historical research (section 2.2.1) and the difficulties when comparing word embeddings among periods of time (section 2.2.2).

2.1 Conceptual History

In the historical field, studying concepts has its own subfield. In intellectual history, the historians who study how concepts have changed over time practice “Begriffsgeschichte” or conceptual history. Influential in conceptual history are the works from

Kosselleck (2002), Foucault (1970) and Skinner (2002). Moreover, Kosselleck (2002) suggests that concepts and their meanings form our political and cultural word view. Therefore, how they change over time are indicators of cultural and political transformations in history (Wevers and Koolen 2020).

Van Sas (1999b) studied different representations of the Dutch nation over the centuries, using political texts and literature from the fifteenth century to 1940. This thesis builds on top of that. With natural language processing methods, we are able to analyze more texts (distant reading) instead of thoroughly analyzing a few selected texts (close reading).

2.2 Word Embeddings

In working with digital sources, methods from computational linguistics can be used to assess how concepts have changed. Word embedding models can be used to trace semantic change over time. Word embeddings are distributional representations of the meaning of words in the context of their neighbors. This means that the meaning of a word can be derived from the context that it appears in (Firth 1957).

Semantical models can be learned, for example with word2vec (Mikolov et al. 2013). With Skip-Gram Negative Sampling (SGNS) models predict contexts or neighboring words within a context window, for a target word. Continuous Bag-of-Words (CBOW), on the other hand, learns by predicting the target words from a given context. In these semantic models, the similarity or dissimilarity of two words is based on the euclidean or cosine distance between vectors (Turney and Pantel 2010).

2.2.1 Word Embeddings for Historical Research. Word embedding models can be used to track semantic change over time. Influential in this field is the study of Hamilton, Leskovec, and Jurafsky (2016). They compared three word embedding models, Positive Pointwise Mutual Information (PPMI), Singular Value Decomposition (SVD) and Skip-Gram Model with Negative Sampling (SGNS), on six corpora, covering over 200 years of data in four different languages. They showed that the word "gay" changed semantically from the meaning of being cheerful to a sexual orientation.

Other studies also used word embeddings in their historical research. Garg et al. (2018) showed that gender and ethnic stereotypes in American books and newspapers from the twentieth and twenty-first century have changed over time. This change is to a similar rate as demographic and occupation shifts of women and ethnic minorities. The same kind of study has been done on Dutch newspapers (Wevers 2019). A more general study on the conceptual change of prominent words in Dutch newspapers between 1950 to 1980 has been done by Orlikowski, Hartung, and Cimiano (2018). Van Lange and Futselaar (2018) have trained word embeddings on Dutch parliamentary proceedings to track how the discussion on the punishment of war criminals changed over time.

2.2.2 The Alignment Problem and Solutions. Comparing word embeddings among different time periods is a difficult task. The numerical representations of words do not have a meaning by themselves. They depend on the relation with other words. The problem is that a word has a different location in one vector than in another. Two word embeddings trained on separate time slices will have two different word embedding spaces, so it is meaningless to compare them. In order to meaningfully compare two word embeddings and their similarities between words, alignment of the word embeddings is necessary. This has also been called the alignment problem (Di Carlo, Bianchi, and Palmonari 2019).

The technique that Hamilton, Leskovec, and Jurafsky (2016) uses is orthogonal procrustes alignment technique, which is done by rotating vectors from all the time slices to be adapted to the first time slice. A downfall to this technique is that the vocabulary across word embeddings for each time slice have to be the same, which in practice means that words that appear not in all time slices have to be pruned. This is a problem for historical research, because language and spelling changes over time (Wevers and Koolen 2020). Moreover, the randomly chosen initialization vectors can have a high influence on the outcome of the word embedding, leading to false conclusions (Hellrich and Hahn 2017; Sommerauer and Fokkens 2019).

Di Carlo, Bianchi, and Palmonari (2019) propose another method: Compass-Aligned Distributional Embeddings. It first trains word embeddings combining all time slices together. These word embedding vectors are then assumed to stay static, so they stay in the same vector space, while context vectors change over time depending on the co-occurrence of words in a time slice. However, this method cannot be used to compare absolute drifts of words between time slices, which is the aim of this thesis.

Because the model of Hamilton, Leskovec, and Jurafsky (2016) requires datasets of 100,000,000 words per time slice, other dynamic word embedding models that work better with sparse data have been proposed. The incremental Skip-Gram model (Kim et al. 2014) relies on the SGNS method. Word vectors of the previous time slice are used to build word vectors for the following time slice. Bamler and Mandt (2017) have created the Dynamic Filtering of Skip-Gram model, which is based on Bayesian word embeddings. In this model, the word vector for the next time slice is learned with a slight drift based on Gaussian variation. Even if the dataset is small, Bayesian inference makes the model robust (Montariol and Allauzen 2019). Yao et al. (2018) uses Positive Pointwise Mutual Information to build their Dynamic-Word2vec model, imposing constraints on sequential vectors, based on the Uhleneck-Ornstein Process.

2.2.3 Dynamic Bernoulli Embedding Model. The model that this research will be using is the Dynamic Bernoulli Embedding model of Rudolph and Blei (2018). This model predicts the distribution of the word embedding, as in CBOW, but is based on a Bernoulli distribution. Rudolph and Blei (2018) have demonstrated that Dynamic Bernoulli Embeddings give better predictive performance for time windows with sparse data. Moreover, this method also captures the change of infrequent words. Both Dynamic Filtering of Skip-Gram and Dynamic Bernoulli Embedding are able to detect drifts with a very sparse dataset, but the Dynamic Bernoulli Embedding has an additional quality to keep words that do not change over time stable (Montariol and Allauzen 2019).

The advantage of the Dynamic Bernoulli Embedding model is, as other word embedding models, that it processes more historical sources at once than traditional historical methods do. In addition, the Dynamic Bernoulli Embedding model also works for more sparse data sources that are often available for historians, in comparison to the large quantity that data science is familiar with. This is also the case for this thesis. Since the earlier periods have smaller amounts of data (see section 3.2), and the three target words in research are not frequently used, Dynamic Bernoulli Embeddings will be effective for this type of research. Additionally, it is important to keep words that do not change over time stable, so that the average drift of words can be meaningfully compared to the absolute drift of the target words.

Rudolph and Blei (2018) use the Dynamic Bernoulli Embedding model to examine words with the highest absolute drift for a specific corpus of interest. They also look at the absolute drift of specific target words that they have chosen. These words are chosen because they expect them to show a significant and interesting semantic shift over time, not because of their topic of research. While this model could contribute to expanding the historical knowledge of ideas, it has not yet been used to expand historical knowledge on a relevant historical discourse. Other kinds of word embeddings have been shown to be practical in research of historical topics such as the position of minorities and the cultural shift in thinking about war crimes (Garg et al. 2018; Wevers 2019; Van Lange and Futselaar 2018). This thesis is the first research that applies the Dynamic Bernoulli Embedding model to Dutch literature to study the semantic shift of words related to nationalism, with the aim of contributing to the historical discourse on nationalism.

3. Data

In this section, the characteristics of the data will be described. Furthermore, the choice of data is justified. In section 3.2 the challenges of working with historical data are discussed.

This study focuses on fiction, because these works were more widespread than non-fiction genres in the DBNL, such as secondary literature and linguistics, as is explained in section 3.1. While rhyme and other stylistic specifics can have an effect of the position and context of the target words, poetry is also included, because it is makes up a large percentage of literary works in the DBNL. This is in particular true for the earlier decades of the time period of interest.

3.1 Historical Literature as Data Source in Historical Context

The Dutch reading public of the early eighteenth century was small. De Kruif (2001) showed that only a quarter of the inhabitants of The Hague owned more than 10 books, and almost 40% of the population did not have any books. Most books in the probate inventories that were used for this study are religious books, such as bibles and hymn books. As time went on, the reading public grew, and inventories showed more varied texts. During the nineteenth century the production of books sped up, costs lowered, and readership expanded. The novel in particular appealed to the middle-class readership (Leerssen 2020b).

Books are seen as one of the main drivers behind nation building (Leerssen 2020b; Anderson 1983; Van den Berg 1999). By reading the same newspapers as well as novels around the nation, "imagined communities" established. An imagined community consists of people of a nation, and while they do not know all members of that community, they feel connected anyway (Anderson 1983). Further, the content of popular genres in the nineteenth century had nationalistic tendencies. The historical novel romantically celebrated the nation's past, while the rustic novel and the realistic novel showed their readers the social and moral representation of the nation (Rigney 2020; Leerssen 2020a). Therefore, literature is a suitable source of data for researching the development of nationalism.

3.2 Challenges

There are two difficulties in working with historical corpora. The first problem with most historical sources is that the further back in time, the less data is available. The reason for this is, that the older a source, the more chance there is that the work has been destroyed (Tosh and Lang 2006). Moreover, it could also be the case that less

Figure 1
Stacked bar chart of the number of tokens before and after preprocessing. Around half of the tokens are removed after preprocessing.

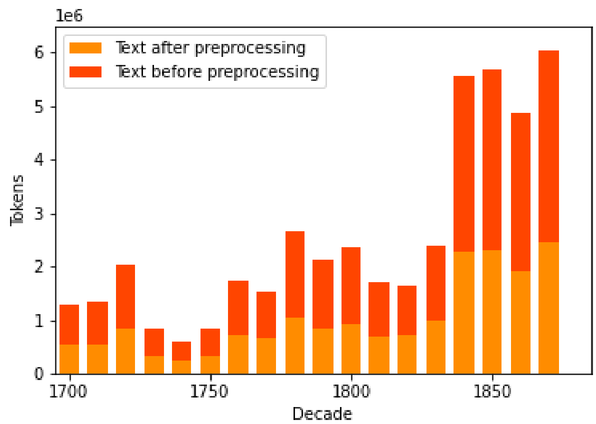
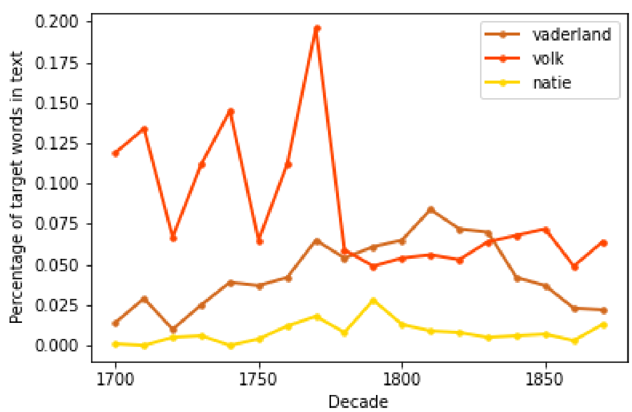


Figure 2
Percentage of occurrences of the target words in the text per decade.

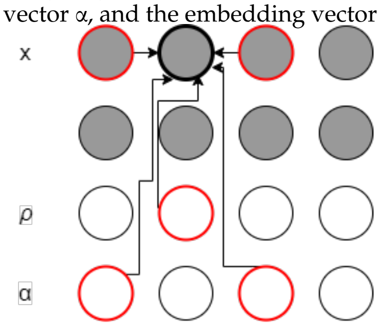


data for earlier periods fits the description "prose", because the novel genre was not as widespread as it is now, as explained above.

For the data this research will be relying on, it is also the case that there is more data for the later periods than there is for the first few, displayed in figure 1. Consequently, this means that the conclusions that can be drawn from the word embeddings for earlier periods are less generalizable as conceptual meanings than those retrieved from the later periods. However, the target words occur in data of all periods. Figure 1 summarizes the occurrences of the target words 'vaderland', 'volk' and 'natie' in the data between 1700 and 1880.

A second challenge in working with historical text data is that aside from the semantic changes that might have occurred, the spelling of words can also have changed over time. Wevers and Koolen (2020) propose two solutions: first, normalize the spelling

Figure 3
Graphical description of the core of the exponential family embedding model, based on the model proposed by Rudolph et al. (2016). The conditional distribution of data point x is governed by the latent context vector α, and the embedding vector ρ.



of words in a preprocessing step, second, correct the spelling after having trained the embeddings. The last step is possible because when there is enough data, words with multiple spellings will be close to each other in the embedding. The method for spelling normalization is explained in the section 4.3.1.

4. Method

In this chapter, the method for this research, Dynamic Bernoulli Embeddings will be explained in section 4.2. Because this method is a type of Exponential Family Embeddings, this class of models is described in section 4.1. After explaining the Dynamic Word Embedding Model, the preprocessing part of the pipeline is described in 4.3. The evaluation methods are explained in section 4.4.

4.1 Exponential Family Embeddings

The idea behind Exponential Family Embeddings of Rudolph et al. (2016) is that each observation depends on the context of other observations it appears in. This class of models is tested with different kinds of analysis, with all sorts of data, such as market basket analysis and neuroscience (Rudolph et al. 2016).

There are three assumptions in Exponential Family Embeddings. First, each data point is conditional on its context, which is described by indices of other data points. Second, this relation can be captured with an appropriate exponential family, or conditional distribution. The two parameters, the latent context vector and the latent embedding vector are combined to form the natural parameter of the family embedding. The third assumption is the embedding structure. The embedding and context vector are shared to learn an embedding of the object of interest. The embedding and context vectors are both learned with the goal of describing features of the data. The objective is the sum of log probabilities of each observation depending on its context, with $L2$ regularization for the embedding and context vectors.[1]

1 The graphical model is based on the graphical model from https://www.youtube.com/watch?v=4s82-SJXhBc&t=81s

4.2 Dynamic Bernoulli Embeddings

Rudolph and Blei (2018) expanded the Exponential Family Embedding to text modeling, with binary observations. Each observation or word x in the text data is indexed, meaning that each word gets an index by position in the vocabulary V, x_{iv}. The word gets an indicator vector with the length of the vocabulary and one non-zero at the position of the word in the vocabulary. Furthermore, each word in the text has its context, c_{i}. x_{ci} are the observations around the data point x.

The conditional distribution of x_{iv} is

$$x_{iv} \mid x_{ci} \sim Bern(p_{iv})$$

where p_{iv} is the Bernoulli probability. The Bernoulli probability of a word in the text (i, v) depends on the embedding vector ρ_{v} and the context vector α_{v}. These vectors depend on terms v and not on positions i, and thus are shared across all positions in the text.

The natural parameter is formed by the inner product of embedding vector ρ_{v} and the context vector α_{v} for words around position i:

$$\eta_{iv} = \rho_v^T (\sum j \in c_i \sum v' \alpha_{v'} x_{jv'})$$

What makes the Dynamic Bernoulli Embeddings dynamic is that each observation x_{iv} is connected to a time slice T_{i}. Then, only the context vectors are shared across all time slices, while the embedding vector differs between time slices.

There is also a prior on the embedding and context vectors, in the form of a Gaussian random walk. This results in more smoothly changing estimates of the embedding of each term:

$$\alpha_v, \rho_v^{(0)} \sim N(0, \lambda_0^{-1} I)$$

$$\rho_v^{(t)} \sim N(\rho_v^{(t-1)}, \lambda^{-1} I)$$

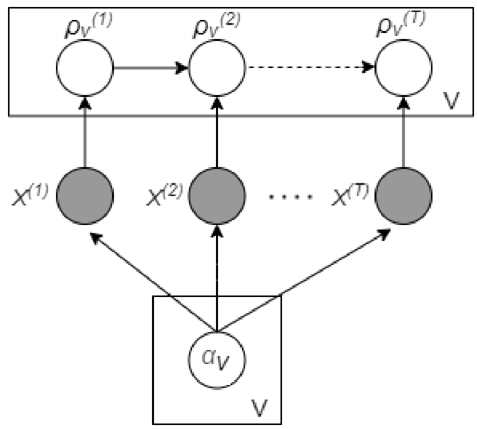
The goal of Dynamic Bernoulli Embeddings and other Exponential Family Embeddings is to learn the embedding vector and the context vector, so that features of the data can be described. In this case, the goal is to find the term embedding for each target word. The context vector and the embedding vectors are fitted with the pseudo log likelihood, which is the sum of log probabilities. The objective for this case is the sum of the log priors and the conditional log likelihoods of the observations, which are divided into non-zero data items L_{pos} and zero data items L_{neg}.

$$L(\rho, \alpha) = L_{pos} + L_{neg} + L_{prior}$$

Where L_{pos} is

$$L_{pos} = \sum_{i=1}^{N} \sum_{v=1}^{V} x_{iv} log \sigma(\eta_{iv})$$

Figure 4
Graphical model of the Dynamic Bernoulli Embedding model, proposed by Rudolph and Blei (2018). The embedding vectors ρ_v change for each time slice $X^{(T)}$, while the context vector α_v is shared.



and L_{neg} is

$$L_{neg} = \sum_{i=1}^{N} \sum_{v=1}^{V} (1 - x_{iv}) log(1 - \sigma(\eta_{iv}))$$

The L_{neg} uses negative sampling.
The prior is

$$L_{prior} = logp(\alpha) + logp(\rho)$$

Where log p(α) is

$$logp(\alpha) = -\frac{\lambda_0}{2} \sum_{v} \|\alpha_v\|^2$$

and log p(ρ) is

$$logp(\rho) = -\frac{\lambda_0}{2} \sum_{v} \left\|\rho_v^{(0)}\right\|^2 - -\frac{\lambda}{2} \sum_{v,t} \left\|\rho_v^{(t)} - \rho_v^{(t-1)}\right\|^2$$

The precision of the random drift, λ, penalizes consecutive word vectors for drifting too far apart. The precision of the random drift enforces smoothness of the word embedding vectors across time slices. The larger the precision of the random drift λ, the more alike the word vectors are over time.

4.3 Preprocessing

In this section, the preprocessing steps will be described. First, in section 4.3, the preprocessing steps that are performed on the raw data, such as spelling normalization (section 4.3.1), stop word removal (section 4.3.3) and dividing the dataset into time slices (section 4.3.4) are outlined. The preprocessing steps are performed with the help the regular expressions library [2] and natural language toolkit (NLTK) (Bird, Loper, and Klein 2009).

4.3.1 Spelling Normalization. The first step of preprocessing is spelling normalization. Spelling normalization is done with rewrite rules. The misspelled parts of words are being replaced with a correctly spelled combination of letters. For example, the "y" in "hy" becomes "ij", so that it the word has the modern spelling "hij" ("he").[3]

The chosen rewrite rules are a combination of three kinds of rules. The first part of rules comes from a previous research on spelling normalization for historical Dutch texts (Braun 2002). Because these rewrite rules are specific to seventeenth century spelling, and most spelling evolves to a modern spelling over time, not all the rules are necessary for eighteenth century and nineteenth century spelling. So, the most appropriate rules for normalizing the text from the centuries of focus in this study are selected.

The second part of rewrite rules are based on systematic errors that can be rewritten with a rule, without modifying correctly spelled words. This requires an adequate balance of specificity and requirement. Not all spelling errors can be corrected, because the list of spelling rules would otherwise become too long, as each misspelled word should have its own individual rewrite rule.[4]

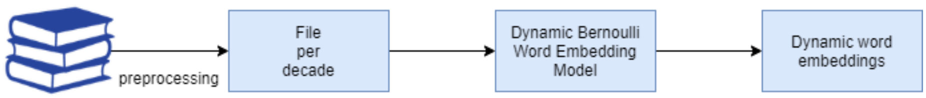
The third part of rules are added after a first round of training. These rules rewrite whole words. The words that need to be rewritten can be divided in two kinds. The first kind of words are frequently misspelled words. These words can be found in the 50 words with the highest absolute drift, so words that do not appear in the same context in the first decades compared to the last decades. Most words in the list have high drift, because the specific spelling is not used anymore in later decades, resulting in that these forms do not appear in later texts. Some of these words in this list are part of the stop words list if they are written according to modern spelling and therefore need to be removed, as explained in section 4.3.3.

The second kind of words that are altered in this last part of rules, are misspelled words that can be found in the nearest neighbors of the target words. It is important to update the spelling of these words, because it might change the order of the nearest neighbors list or even the words in the nearest neighbors list themselves.

4.3.2 Stemming. In addition, stemming with the Kraaij Pohlmann stemmer[5] was tried as a means to lower spelling variation. Especially for spelling of plural forms, this could be beneficial. When words end in one consonant and have a double vowel in the final

2 `https://docs.python.org/3/library/re.html`
3 The rewrite rules can be found in the code.
4 For example, Dutch modern spelling barely uses the character "y", so one of the rewrite rules is that "y" becomes "ij". However, this means that some words become incorrect, in this case, the word "Egypte" ("Egypt") becomes "Egijpte". This choice can be justified by the fact that words like "Egypt" are rarely used in these texts in comparison to other words in Dutch language with an "ij", such as the word "he".
5 `http://snowball.tartarus.org/algorithms/kraaij_pohlmann/stemmer.html`

Figure 5
Graphical overview of the pipeline used



syllable of the singular form, the double vowel will become one vowel in plural form. By stemming all the forms are reduced to the singular modern form.[6]

However, because language has so many variations, the Kraaij-Pohlman stemmer is not perfect and could result in even more spelling errors. Furthermore, many words are unrecognizable after stemming, which is undesirable when the goal of this research is showing and understanding in which context the target words are used.

4.3.3 Stop Word Removal. Removing most frequent words is done by removing all Dutch stop words assigned by the NLTK package. Because the most frequently used words in historical texts might be different than in modern language, the most frequently used words from multiple texts from 1700 to 1750 and multiple texts from 1840-1880 are added to the list of stop words.

There is an additional pruning of words to make sure that words are at least two characters long, and don't consist of numbers. Moreover, during the creation of the dictionary, a mapping between a term and its integer id, words that are in less than 10 of the documents are filtered. This results in a compacter dictionary for the model.

4.3.4 Time Slicing. For each book in the dataset, a file is created in which the preprocessed text is stored. These texts have the aforementioned normalized spelling, and are without stop words, punctuation and numbers. Then, the preprocessed documents from the same decade are concatenated to one file. In total, eighteen text files are created, one for each time slice.

4.4 Evaluation

Two methods are useful for evaluating dynamic word embeddings. The first method is testing the Bernoulli probability of words in the test set. Rudolph and Blei (2018) use the L_{pos} value as metric, which shows how good the model predicts the target word from given the context from the validation set.

A second way of evaluating dynamic word embedding models is by using control words. Control words are concepts that do not have a relation with the topic. If the word embeddings for the target words change over time, the word embedding of a neutral control word should not reveal a similar change (Sommerauer and Fokkens 2019). This evaluation method could be seen as a robustness check of the model.

To see if words of different parts of speech behave differently, three control words are chosen. A verb: "lopen" ("walk"), a noun: "vis" ("fish"), and a more abstract noun:

6 An example: "traan" ("tear") becomes "tranen" ("tears"). However, in earlier Dutch spellings, it would also be acceptable to write "traanen" as a plural form. The Kraaij-Pohlman Dutch stemmer stems plural forms as a singular form. For example, instead of just removing the plural -en ("tranen" becomes "tran"), it recognizes that "tranen" should become "traan". This results in all the forms of "traan", whether plural or singular, modern or early spelling, are limited to the singular modern form.

Table 1
Hyperparameter settings

	values
minibatches	100, 300 and 500
learning rate	0.2, 0.02, 0.002 and 0.0002
drift	1, 5 and 10

"liefde" ("love"). Because the three target words are all rather abstract, they might have a more substantial change over time than non-abstract words.

5. Experimental Setup

The text data is divided in a training dataset and a validation dataset, in a ratio of 80% and 20%. After 100 mini batches, the positive likelihood (L_{pos}) is calculated on the validation set and saved.

The number of passes over the data is ten, with an additional first pass, or zeroth pass, where the embedding vector is trained on all the time slices, for initialization. The dimension of the embeddings are set to 100, and the number of negative samples is set to 20. These settings are based on the settings of Rudolph and Blei (2018).

5.1 Context size

An important setting of a word embedding model is the context size. A smaller context size results in nearest neighbors that are of the same syntactic category. If a larger context size is used, the nearest neighbors of target words are topically related instead (Levy and Goldberg 2014). Because this thesis focuses more on the broader context in which the three target words are used, larger context windows are required.

Moreover, the right size of a context window is language dependent. A context window of five is often chosen to study topical similarity in English (Levy and Goldberg 2014), while studying conceptual association with Dutch text requires a somewhat larger window (Wevers and Koolen 2020). Therefore, a context size of six is chosen.

5.2 Hyperparameter Settings

The model has three hyperparameters that are needed to be set: the batchsize, the learning rate and the precision of the random drift. The choice of the possible hyperparameter settings are based on the settings of Rudolph and Blei (2018) and are around the default settings of the model.

The model is expensive to run (6 to 7 hours on the server of Google Colab), so for efficiency reasons, instead of testing a combination of all the settings, we first test the best batch size setting, while keeping the other hyperparameters on the default setting. We keep the batch size setting that gives the highest L_{pos} on the evaluation set, for comparing different learning rates. This is repeated for the last setting, the precision of the random drift. Then, the model with the one with the highest L_{pos} is chosen to be the final model.

Table 2
Results for hyperparameter setting

Minibatches	Learning rate	Drift	Lpos val
100	0.002	1	-6854684
300	0.2	1	-11756400
	0.02	1	-6560637
	0.002	1	-6281450
	0.002	5	-7143437
	0.002	**10**	**-6064922**
	0.0002	1	-7961574
500	0.002	1	-6287337

6. Results

In this chapter, the results will be described. The results for the hyperparameter tuning will be outlined in section 6.1. Then, the results of the absolute drift of the target words will reported in section 6.2 and the results of the nearest neighbors of the target words in section 6.3.

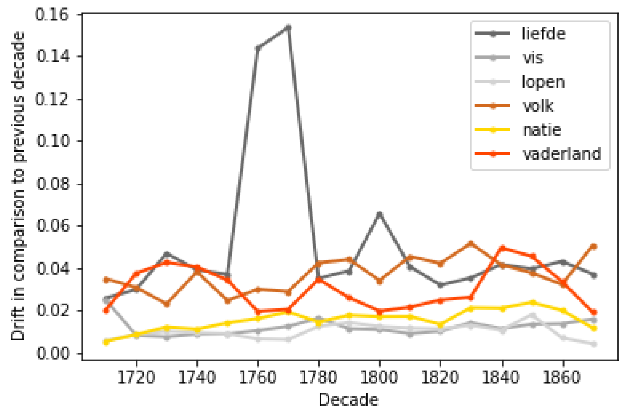
6.1 Evaluation of the Word Embeddings

As explained in section 4.4, Dynamic Bernoulli Embedding models are evaluated with the Bernoulli positive likelihood on the validation set, or L_{pos}. This metric is used to select the hyperparameter settings, such as the minibatch size, the learning rate and the precision of the random drift. As explained in section 5.2 we first train with the different settings of the minibatch size, holding all other hyperparameters constant. After finding the best minibatch size, we explore the settings of the learning rate. Then, the best precision of the random drift is searched for, while selecting the best minibatch setting and the best learning rate.

The results of the experiments are represented in table 2. The setting that receive optimal results on the held-out validation set is the model with 300 minibatches, a learning rate of 0.002 and a precision of the random drift of 10. However, for the first hyperparameter setting that was looked at, the number of minibatches, the L_{pos} for 300 and 500 minibatches, are very close to each other.

In addition to looking at the highest L_{pos} as evaluation method, neutral control words are being used to see if the semantic drift of the target words behaves by coincidence and resembles other words not related to the topic (Sommerauer and Fokkens 2019). As depicted in figure 6.2 all the words have their own drifting pace. As expected, the noun "vis" ("fish") and the verb "lopen" ("walk"), did not semantically change over time, the drift remained rather steady. More abstract words like "liefde" ("love") and "vaderland" ("fatherland") have a larger drift.

Figure 6
The drift of target words and neutral words, in comparison to their position in the previous decade



6.2 Absolute Drift

The absolute drift is the metric used to measure how much the context, or the usage of a word changes over time. Table 3 shows the absolute drifts for the target words. The mean absolute drift statistic is mean absolute drift for all the tokens that are left in the text files after filtering. This statistic will be used as a baseline.

For the final model with drift precision of the random drift of 10, the target word that changes the most over time is "volk", followed by "vaderland". The target word "natie" changed the least over time. All three words have a higher absolute drift than the mean absolute drift.

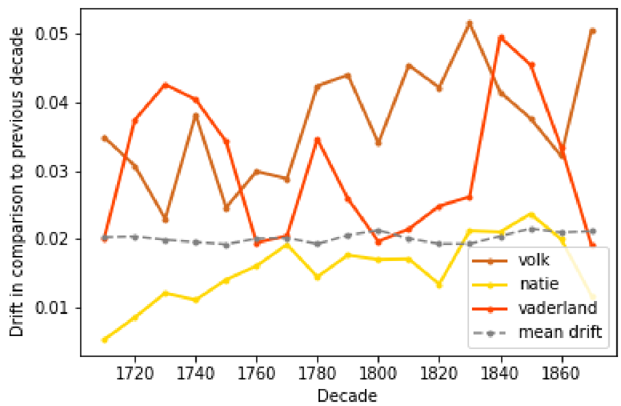
To give another frame of reference, the word of the dataset that changed the most was "we" (informal form of "wij", "we"), with an absolute drift of 0.3864. Looking at the position of the words with the largest absolute drift, the words "volk", "vaderland", "natie" are in the 157th, 882nd and 3711th place of the 61114 terms respectively.[7]

Table 3
Absolute drifts for the target words, for the model with drift 1 and drift 10

Target Word	Embedding Drift 1	Embedding Drift 10
Mean absolute drift	0.0982	0.0253
"natie"	0.2641	0.0644
"volk"	0.4050	0.1800
"vaderland"	0.4310	0.1128

7 For a list of the 25 words with the largest drift, see Appendix A

Figure 7
The drift of only the target words in comparison to their position in the previous decade and the average drift of all the words in comparison to the previous decade



6.2.1 Drift of Target Words over Time. Figure 7 shows the drift of target words over time. This graph shows that while "natie" has a small absolute drift, the drift becomes a bit larger over time, but it stays below the average drift of words. The graph for "vaderland" shows three peaks: during the first half of the 18th century, the last two decades of the eighteenth century and around the middle of the nineteenth century. "Volk", the target word with the largest drift, shows a constant change in the context of the word. However, around the turn of the nineteenth century and 1870 the drift is the largest.

6.3 Nearest Neighbors of Target Words

The nearest neighbors are the words that are closest to the target word in the embedding at a specific time slice. The nearest neighbors can be understood as the word most often used in a similar context of the target words, and thus are most like the context word.

In table 4, the five nearest neighbors of the target word "vaderland" are listed. The table shows small changes for the nearest neighbors over 18 decades. Almost all words remain in the top 5 nearest neighbors, while only one new word replaces a nearest neighbor. Also, the position of the words in the top 5 nearest neighbors shows minimal variation over time. This is also the case for the five nearest neighbors of the target words "volk" and "natie". The nearest neighbors on place 6 to 20 vary more over time. The results for the other target words and with a larger list of nearest neighbors are found in Appendix A.

7. Discussion

In this chapter the findings of chapter 6 will be discussed. In section 7.1, the results of the terms with the highest absolute drift in relation to the use of historical fiction literature will be discussed. In section 7.2 the results of the nearest neighbors of the target words will be presented. These findings will be compared with the results of the second best model with a smaller precision of the drift prior. The chapter ends with section 7.3: an analysis of how the nearest neighbors of the target words “natie”, “volk” and “vaderland” have changed over time and how this links to the development of nationalism in the eighteenth and nineteenth century.

Table 4
The 5 nearest neighbors for the target word "vaderland" over the decades

1700-1710	1740-1750	1790-1800	1830-1840	1870-1880
geboorteland	geboortegrond	geboorteland	geboorteland	geboorteland
geboortegrond	geboorteland	volksbestaan	geboortegrond	eendrachtsband
volksbestaan	volksbestaan	geboortegrond	volksbestaan	dierbaar
dierbaarst	dierbaarst	dierbaarst	dierbaar	geboortegrond
eendrachtsband	eendrachtsband	eendrachtsband	eendrachtsband	dierbaarst

7.1 Absolute Drift

When looking at the terms with the largest drift[8], two observations stand out. Firstly, some of the words that have the largest drift are words with old spelling forms or are old-fashioned words. They appear in the words with the highest absolute drift because these terms are not used anymore in later decades, and therefore their position in the word embedding only relies on the drifting prior mechanism.

The presence of words with old spelling forms shows that not all spelling is corrected in the spelling normalization step, even though a list of words that are frequently misspelled, with their modern spelling is added to the rewrite rules in the preprocessing step. This means that the word embeddings could be improved by further spelling normalization.

Secondly, some of the other words that have the largest drift are names. This is the result of using fictional literature with fictional characters. As with the older spelling forms, names of characters only appear in some books, in some decades. This means that for some decades the place of a name in the embedding depends entirely on the drifting prior mechanism.

7.1.1 Drift of Target Words over Time. As explained in section 6.2.1, the word "vaderland" shows three peaks in their drift over time. The first peak coincides roughly with the emergence of the word "vaderland" in Dutch book titles (Kloek 1999). The second peak is around 1780, which is the decade of the politicization of the Dutch enlightenment (Kloek 1999). The third peak of drift happens in the decade of the Dutch constitutional reform of 1848 and the revolutions in Europe of the same year.

The word "volk" changes fast from the last quarter of the eighteenth century onwards. This also the time of politicization of the Dutch Enlightenment. As will be explained in section 7.3.2, the nearest neighbors of "volk" in the earlier decades of the eighteenth century are mainly related to biblical themes. The relation to people of Israel changes more to the interpretation as people as seen as mob in later decades, which might explain the larger drift from the 1780s onwards.

8 For a list with words with the highest absolute drift, see Appendix C

Table 5
The 5 nearest neighbors for the target word "vaderland" over the decades with a precision drift of 1

1700-1710	1740-1750	1790-1800	1830-1840	1870-1880
gebroortegrond	geboortegrond	geboortegrond	geboortegrond	geboortegrond
land	volksbestaan	volksbestaan	bataven	vlaanderland
volksbestaan	onafhankelijkheid	land	nederland	bataven
behulpzame	bataven	nederland	volksbestaan	eendrachtsband
getuchtigd	volksgeest	gemenebest	land	vrijgevochten

7.2 Nearest Neighbors

As the nearest neighbors and the absolute drift shows (table 4 and table 3), the semantic change of the target words is not that large. While it is possible that the context that these words are used in do hardly evolve, on the other hand, these small changes could also be the consequence of the large precision of the random drift. This penalizes sequential word embedding vectors that drift too far apart.

To investigate this option, we can compare the nearest neighbors of the final model with the second-best performing model, that has the same hyperparameter settings except for a precision of the random drift of 1, instead of 10.[9] The table of nearest neighbors of the target word "vaderland" in multiple decades are displayed in table 5.[10] From this table it can be noticed that while some words as "geboortegrond" ("birthplace") are relatively stable over time, other words appear and disappear in the top five, depending on the decade.

As can be read in the next section, a lesser restriction can benefit interpretation for historical research, even when it is advised to choose the model with the precision of the random drift of 10 as final model, relying on the L_{pos} score on the validation set.

7.3 Interpretation of Nearest Neighbors of the Target Words

In this section the nearest neighbors of the target words, and how they have changed over time will be discussed. The nearest neighbors are limited to the top 10 nearest neighbors for the target words. The top 20 can be found in Appendix A. In Appendix B the results of the 20 nearest neighbors from the model with precision drift 1 are presented.

7.3.1 Vaderland. The words that are most similar to "vaderland" are the words "geboorteland" ("country of birth") and "geboortegrond" ("place of birth"), which are indeed very similar to the word vaderland. There are also some affectionate words such as "dierbare" and "dierbaarst" ("(most) precious") and "vrijgevochten" ("free-spirited"). Moreover, in the top 10 also words that relate to a fatherland's past appear,

9 For the drifts of target and control words for the model with a precision drift of 1, see Appendix D
10 The tables with a longer list of nearest neighbors can be found in Appendix A, as well as the list of nearest neighbors for other target words.

such as “wapenroem” ("fame of arms"), “onafhankelijkheid” ("independence")[11], and especially related to the Dutch past: “bataven”. Furthermore, the words “nederland” ("The Netherlands") and “vlaanderland” ("Flanders") are present in the top 10 most similar words from 1780s onwards.

The results for nearest neighbors for “vaderland” with a smaller precision of the random drift shows the same or similar words. What is different, however, is that these results also show the offset of the Belgian independence. From the 1820s onwards “vlaanderland” (roughly translated "Flanderland") appears in the top 10, just a few years away from the Belgian Revolution of 1830. Moreover, “vlaanderland” rises in similarity over the decades and becomes the second closest word to “vaderland” in the 1870s.

7.3.2 Volk. When the nearest neighbors for the target word “volk” are looked at, the results from the final model and the model with precision of the random drift of 1 are fairly similar. The nearest neighbor of “volk” is “oproerig” ("rebellious") for every decade except the last one. Other words like “oproer” ("rebellion"), “muitziek” and “muitzucht” ("mutinious") emphasize the dangerous side of the people, in the meaning of people as a mob. These words appear more frequently and get a higher position in the top 10 nearest neighbors in the later decades.

The nearest neighbors of the earlier decades show the result of the biblical literature and themes that were popular in the eighteenth century (De Kruif 2001). The terms that are related to the word “volk” in these decades are terms related to the people of the Bible, such as “isreal”, “israal”, “juda”, "egipten" ("Egypt"), “sanhedrin" and the people of Amalek (“amalek”).

Words that could have a connection to the meaning of “people” as in the inhabitants of a nation are “landzaat” ("inhabitant of a country") and “vrijheidlievend” ("freedom-loving"). They appear in last decades. In the model with precision of the random drift of 1, the words “natie” ("nation") and “burgerrecht” ("civil right") are present in the nearest neighbors in 1870s.

7.3.3 Natie. While the outcomes of the final model with a precision of the random drift of 10 does not show a traceable semantic drift of the word “natie”, it is interesting to look at the semantic similar words to “natie”. “Handeldrijvende” ("trading") is most similar in all decades, “naäpen” ("copying", as in what a copycat does) is second most similar. This means that trading is often spoken of in the context of nations, and also nations copying each other. Also, nationality (“nationaliteit”) is third most similar. In addition, institutions are mentioned, such as universities (“universiteiten”), courts (“gerechtshoven”, “rechtbanken”), governments (“gouvernments”), people’s government (“volksregering”), and republic (“republiek”). The last two terms, could be related to the Dutch Republic.

The second-best model with a smaller prior on the drift shows more variation. In the 1790s, the nearest neighbor of "natie" is “nederlanders” (the Dutch people), which coincides with the emergence of the Dutch Batavian Republic (Kloek and Mijnhardt 2001). The Dutch Constitution of 1848 did also leave its mark on the word embeddings: the term "constitution" appears from 1840s onward in the top 10. Also, the word “vrijheidlievend” (“freedom-loving”), which might be a national Dutch or Belgian characteristic, gets a higher position over time.

11 The word independence could also relate to the future, in terms of the Belgian independence

8. Conclusion

In this thesis the emergence and development of nationalism in Dutch culture was researched by looking at how the context of the target words "natie", "volk" and "vaderland" have changed. This has been done by applying the Dynamic Bernoulli Model, proposed by Rudolph and Blei (2018) to Dutch fiction literature between 1700-1880. This period has been chosen because it roughly corresponds to the Sattelzeit in which society developed from the early modern period to the modern period (Brunner et al. 1972). This also meant the emergence and development of nation building and nationalism (Gellner 1983). To our knowledge this is the first research that uses Dynamic Bernoulli Embeddings to contribute to a historical discourse, which is in this case the debate on the origins and spread of nationalism.

The best results in terms of intrinsic evaluation were achieved with a model that heavily penalizes word vector embeddings for drifting too far apart. This resulted in the absolute drift of words to be low, and that the nearest neighbors of the target words did not change much over time. This model was compared to a model with a less severe restriction on the random drift precision. The second model was able to show the political and cultural shifts in Belgian and Dutch societies, and therefore, may be more suitable to be used in historical research.

Analysis of word embeddings showed that the word "natie" did not undergo a large semantic shift over time, while the words "vaderland" and "volk" showed a slightly larger semantic shift.

This was also noticeable in the nearest neighbors, the words that appear in the same context as the context of the target words. "Volk" shifted from the biblical context to the context in which the people were seen as a "mob". "Vaderland" had since the early decades of research a nationalistic context, with words that glorify the nation's past, and positive national characterizations. Later decades show a greater prominence of the words "nederland" and "vlaanderland". "Natie" showed almost no variation over time, as was expected by the low absolute drift. Words that are used in the same context as "natie" are related to institutions. In contrast, the results of the nearest neighbors in the second model, the word "constitution" appeared in the top 10 nearest neighbors for the word "natie", just before the declaration of the Dutch constitution.

8.1 Future Research

A start has been made to map the semantic shifts of the words "natie", "volk" and "vaderland" in Dutch fiction literature of the eighteenth and nineteenth century. An in depth analysis of how the semantic shifts of the target words reflect the development of nation building and nationalism is beyond the scope of this thesis. The results from the word embeddings of this thesis could be a starting point of further research on nationalism.

In the short time period of seven weeks, not all hyperparameter settings could be tested. In future work, the settings of hyperparameters could be improved by testing other combinations as well, especially since L_{pos} outcomes were very close to each other and the random drift precision seemed to have a strong influence on the validation metric. Moreover, as Hellrich and Hahn (2017) emphasize, the random initialization influences the outcomes of the model too.

Another interesting topic for future research is to see how lemmatization or stemming influences the outcomes of the models and how useful they might by for researching historical texts.

References

Anderson, Benedict R. O'G. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso, London.

Bamler, Robert and Stephan Mandt. 2017. Dynamic word embeddings. In *Proceedings of the 34th International Conference on Machine Learning, volume 70 of Proceedings of Machine Learning Research*, pages 380–389, PMLR, Sydney, Australia.

Van den Berg, Willem. 1999. Verbeelding van het vaderland. Het denken over het vaderland in de letterkunde van de eerste decennia van de negentiende eeuw. In Niek van Sas, editor, *Vaderland*, number I in Nederlandse Begripsgeschiedenis. Amsterdam University Press, Amsterdam.

Bird, Steven, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*. O'Reilly Media, Beijing.

Braun, Loes. 2002. Information Retrieval from Dutch Historical Corpora. Master's thesis, Maastricht University.

Brunner, Otto, Werner Conze, Reinhart Koselleck, and Arbeitskreis für Moderne Sozialgeschichte, editors. 1972. *Geschichtliche Grundbegriffe*, 4. aufl edition. Number historisches Lexikon zur politisch-sozialen Sprache in Deutschland / hrsg. von Otto Brunner; Werner Conze; Reinhart Koselleck. [Hrsg. im Auftrag des Arbeitskreises für Moderne Sozialgeschichte e.V.] ; Bd. 2 in Geschichtliche Grundbegriffe. Klett-Cotta, Stuttgart. OCLC: 246138897.

De Kruif, José. 2001. Classes of readers: Owners of books in 18th-century The Hague. *Poetics*, 28(5-6):423–453.

Di Carlo, Valerio, Federico Bianchi, and Matteo Palmonari. 2019. Training Temporal Word Embeddings with a Compass. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6326–6334.

Firth, J.R. 1957. A synopsis of linguistic theory 1930-1955. In Frank Palmer, editor, *Selected Papers of J.R. Firth 1952-1959*. Longman.

Foucault, Michel. 1970. The archaeology of knowledge. *Social Science Information*, 9(1):175–185.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644.

Gellner, Ernest. 1983. *Nations and nationalism*. New perspectives on the past. Blackwell, Oxford.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 2016. 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, Association for Computational Linguistics.

Hellrich, Johannes and Udo Hahn. 2017. Don't Get Fooled by Word Embeddings-Better Watch their Neighborhood. In *DH*.

Kim, Yoon, 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, Association for Computational Linguistics, Baltimore, MD, USA.

Kloek, J.J. 1999. Vaderland en letterkunde. In Niek van Sas, editor, *Vaderland*, number I in Nederlandse Begripsgeschiedenis. Amsterdam University Press, Amsterdam.

Kloek, J.J. and W.W. Mijnhardt. 2001. *1800. Blauwdrukken voor een samenleving*. Number 5 in 'De Nederlandse cultuur in Europese context'. Sdu Uitgevers.

Kosselleck, R. 2002. *The practice of conceptual history: Timing history, spacing concepts.* Stanford University Press.

Leerssen, J.T. 2020a. Literary realism and the nation. In J.T. Leerssen, editor, *Encyclopedia of Romantic Nationalism in Europe*. Amsterdam University Press.

Leerssen, J.T. 2020b. Publishing, libraries, book clubs and reading societies. In J.T. Leerssen, editor, *Encyclopedia of Romantic Nationalism in Europe*. Amsterdam University Press.

Levy, Omer and Yoav Goldberg. 2014. Dependency-Based Word Embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 302–308, Association for Computational Linguistics, Baltimore, Maryland.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in neural information processing systems*, volume 26, pages 3111–3119, Curran Associates, Inc.

Montariol, Syrielle and Alexandre Allauzen. 2019. Empirical Study of Diachronic Word Embeddings for Scarce Data. In *Proceedings of Recent Advances in Natural Language Processing*, pages 795–803, Varna, Bulgaria.
Orlikowski, Matthias, Matthias Hartung, and Phillipp Cimiano. 2018. Learning Diachronic Analogies to Analyze Concept Change. In *Conference: 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 1–11, Santa Fe, New Mexico, United States of America.
Rigney, Ann. 2020. The historical novel. In J.T. Leerssen, editor, *Encyclopedia of Romantic Nationalism in Europe*. Amsterdam University Press.
Rudolph, Maja and David Blei. 2018. Dynamic Embeddings for Language Evolution. In *WWW*, pages 1003–1011.
Rudolph, Maja, Francisco Ruiz, Stephan Mandt, and David Blei. 2016. Exponential family embeddings. In *Advances in Neural Information Processing Systems*, pages 478–486.
Skinner, Quentin. 2002. *Visions of politics*. Cambridge University Press.
Sommerauer, P. and A. Fokkens. 2019. Conceptual change and distributional semantic models: An exploratory study on pitfalls and possibilities. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 223–233, Association for Computational Linguistics (ACL).
Tosh, J. and S. Lang. 2006. *The pursuit of history: Aims, methods and new directions in the study of modern history*. Pearson Education.
Turney, P. D. and P. Pantel. 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37:141–188.
Van Lange, M. and R. Futselaar. 2018. Debating evil: Using word embeddings to analyze parliamentary debates on war criminals in The Netherlands. In *Proceedings of the Conference on Language Technologies & Digital Humanities*, pages 147–153, Znanstvena založba Filozofske fakultete v Ljubljani.
Van Sas, Niek. 1999a. De vaderlandse imperatief. Begripsverandering en politieke conjunctuur, 1763-1813. In Niek van Sas, editor, *Vaderland*, number I in Nederlandse Begripsgeschiedenis. Amsterdam University Press, Amsterdam.
Van Sas, Niek, editor. 1999b. *Vaderland. Een geschiedenis vanaf de vijftiende eeuw tot 1940*. Amsterdam University Press.
Wevers, M. 2019. Using word embeddings to examine gender bias in Dutch newspapers, 1950-1990. pages 92–97.
Wevers, M. and M. Koolen. 2020. Digital begriffsgeschichte: Tracing semantic change using word embeddings. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(4).
Yao, Zijun, Yifan Sun, Weicong Ding, Nikhil Rao, and Hui Xiong. 2018. Dynamic Word Embeddings for Evolving Semantic Discovery. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM '18, pages 673–681, Association for Computing Machinery, New York, NY, USA.

Appendix

A. Top 20 Nearest Neighbors for the model with precision of the random drift of 10

Table 6
Nearest neighbors for the target word "vaderland" from the model with drift 10 (1/3)

1700	1710	1720	1730	1740	1750
geboorteland	geboorteland	geboortegrond	geboortegrond	geboortegrond	geboorteland
geboortegrond	geboortegrond	geboorteland	volksbestaan	geboorteland	volksbestaan
volksbestaan	volksbestaan	volksbestaan	geboorteland	volksbestaan	geboortegrond
dierbaarst	dierbaarst	dierbaarst	dierbaarst	dierbaarst	dierbaarst
eendrachtsband	eendrachtsband	eendrachtsband	dierbaar	eendrachtsband	eendrachtsband
onafhanklijkheid	onafhanklijkheid	onafhanklijkheid	eendrachtsband	onafhanklijkheid	vrijgevochten
bataven	bataven	bataven	onafhanklijkheid	dierbaar	onafhanklijkheid
wapenroem	wapenroem	dierbaar	bataven	vrijgevochten	bataven
dierbaar	vrijgevochten	vrijgevochten	vrijgevochten	bataven	wapenroem
banneling	dierbaar	wapenroem	wapenroem	wapenroem	vaderlande
vrijgevochten	vaderlande	nederland	nederland	nederland	dierbaar
vaderlande	roemvol	roemvol	slavenband	roemvol	roemvol
roemvol	banneling	slavenband	roemvol	vaderlande	nederland
burgeres	burgeres	nassaus	vaderlande	slavenband	vlaanderland
vlaanderland	nassaus	vaderlande	nassaus	nassaus	banneling
nassaus	slavenband	banneling	vlaanderland	vlaanderland	nassaus
nederland	vlaanderland	vlaanderland	oorlogsroem	oorlogsroem	burgeres
land	nederland	oorlogsroem	banneling	banneling	oorlogsroem
slavenband	volksgeluk	staatsgebouw	staatsgebouw	staatsgebouw	slavenband
slavenband	volksgeluk	staatsgebouw	staatsgebouw	staatsgebouw	slavenband

Table 7
Nearest neighbors for the target word "vaderland" from the model with drift 10 (2/3)

1760	1770	1780	1790	1800	1810
geboorteland	geboorteland	geboorteland	geboorteland	geboorteland	geboorteland
geboortegrond	volksbestaan	volksbestaan	volksbestaan	geboortegrond	geboortegrond
volksbestaan	geboortegrond	geboortegrond	geboortegrond	volksbestaan	volksbestaan
dierbaarst	dierbaarst	dierbaarst	dierbaarst	dierbaarst	dierbaar
eendrachtsband	eendrachtsband	eendrachtsband	eendrachtsband	dierbaar	dierbaarst
vrijgevochten	onafhanklijkheid	vrijgevochten	dierbaar	eendrachtsband	eendrachtsband
onafhanklijkheid	vrijgevochten	dierbaar	onafhanklijkheid	onafhanklijkheid	onafhanklijkheid
wapenroem	wapenroem	bataven	vrijgevochten	vaderlande	vaderlande
vaderlande	bataven	onafhanklijkheid	bataven	bataven	vrijgevochten
roemvol	dierbaar	vlaanderland	vlaanderland	vlaanderland	nederland
nederland	vaderlande	nederland	vaderlande	nederland	bataven
dierbaar	nederland	wapenroem	wapenroem	vrijgevochten	wapenroem
bataven	roemvol	vaderlande	nederland	wapenroem	vlaanderland
vlaanderland	vlaanderland	roemvol	roemvol	roemvol	roemvol
banneling	slavenband	nassaus	nassaus	banneling	banneling
burgeres	banneling	slavenband	banneling	burgeres	nassaus
volksgeluk	oorlogsroem	oorlogsroem	oorlogsroem	nassaus	volksgeluk
nassaus	nassaus	leeuwenmoed	slavenband	oorlogsroem	oorlogsroem
oorlogsroem	burgeres	banneling	burgeres	volksgeluk	slavenband
slavenband	volksgeluk	neerland	volksgeluk	land	burgeres

Table 8
Nearest neighbors for the target word "vaderland" from the model with drift 10 (3/3)

1820	1830	1840	1850	1860	1870
geboorteland	geboorteland	geboorteland	geboorteland	geboorteland	geboorteland
geboortegrond	geboortegrond	volksbestaan	eendrachtsband	eendrachtsband	eendrachtsband
volksbestaan	volksbestaan	geboortegrond	volksbestaan	geboortegrond	dierbaar
dierbaar	dierbaar	eendrachtsband	geboortegrond	volksbestaan	geboortegrond
dierbaarst	eendrachtsband	dierbaar	dierbaar	dierbaar	dierbaarst
eendrachtsband	dierbaarst	dierbaarst	bataven	dierbaarst	volksbestaan
vaderlande	vrijgevochten	bataven	vlaanderland	bataven	bataven
onafhanklijkheid	bataven	vrijgevochten	dierbaarst	vlaanderland	vlaanderland
vrijgevochten	vaderlande	nederland	vrijgevochten	vrijgevochten	vrijgevochten
bataven	nederland	onafhanklijkheid	onafhanklijkheid	onafhanklijkheid	slavenband
nederland	vlaanderland	vlaanderland	wapenroem	slavenband	onafhanklijkheid
vlaanderland	onafhanklijkheid	vaderlande	land	wapenroem	wapenroem
wapenroem	land	land	neerland	vaderlande	vaderlande
roemvol	wapenroem	wapenroem	vaderlande	nassaus	nassaus
nassaus	roemvol	nassaus	nassaus	heldenkroost	heldenkroost
slavenband	nassaus	slavenband	slavenband	neerland	burgertrouw
volksgeluk	oorlogsroem	heldenkroost	nederland	burgertrouw	neerland
oorlogsroem	slavenband	burgertrouw	heldenkroost	land	land
banneling	volksgeluk	neerland	burgertrouw	oorlogsroem	oorlogsroem
belgie	belgie	oorlogsroem	oorlogsroem	dwangzucht	dwangzucht

Table 9
Nearest neighbors for the target word "volk" from the model with drift 10 (1/3)

1700	1710	1720	1730	1740	1750
oproerig	oproerig	oproerig	oproerig	oproerig	oproerig
israël	israël	israël	israël	israël	israël
oproer	oproer	oproer	volks	volks	volks
amalek	amalek	volks	oproer	oproer	oproer
volks	volks	amalek	amalek	amalek	koningdom
juda	juda	juda	koningdom	koningdom	amalek
heidnen	koningdom	koningdom	juda	juda	sanhedrin
israal	heidnen	heidnen	heidnen	muitziek	juda
koningdom	muitend	sanhedrin	muitziek	sanhedrin	muitziek
sanhedrin	muitziek	muitziek	sanhedrin	oproers	heidnen
isrels	muitzucht	isrels	muitend	muitend	oproers
isrel	sanhedrin	muitend	isrels	heidnen	heidendom
muitziek	isrels	israal	muitzucht	muitzucht	muitend
muitend	oproers	heidendom	isrel	oproerigeid	oproerigeid
heidendom	isrel	isrel	oproers	vloekverbond	muitzucht
oproers	vloekverbond	oproers	heidendom	heidendom	handeldrijvend
muitzucht	vloekverwanten	muitzucht	israal	handeldrijvend	souvereinen
vloekverwanten	jakobs	jakobs	vloekverbond	landzaat	vloekverbond
vloekverbond	israal	vloekverbond	jakobs	isrels	israal
jakobs	heidendom	vloekverwanten	landzaat	souvereinen	landzaat

Table 10
Nearest neighbors for the target word "volk" from the model with drift 10 (2/3)

1760	1770	1780	1790	1800	1810
oproerig	oproerig	oproerig	oproerig	oproerig	oproerig
israël	israël	oproer	oproer	oproer	volks
oproer	volks	israël	israël	volks	oproer
volks	oproer	volks	volks	israël	israël
koningdom	koningdom	koningdom	koningdom	koningdom	koningdom
amalek	amalek	sanhedrin	amalek	amalek	muitzucht
juda	muitziek	amalek	oproers	muitzucht	amalek
sanhedrin	sanhedrin	oproers	muitzucht	oproers	muitziek
muitziek	muitzucht	muitziek	handeldrijvend	muitziek	muitend
heidnen	muitend	muitzucht	verjaagde	handeldrijvend	oproers
muitend	juda	muitend	muitend	muitend	landzaat
heidendom	oproers	vloekverbond	muitziek	souvereinen	vloekverbond
oproers	vloekverbond	heidnen	souvereinen	sanhedrin	oproerigeid
muitzucht	landzaat	juda	sanhedrin	volksregering	sanhedrin
israal	heidnen	souvereinen	oproerigeid	oproerigeid	handeldrijvend
vloekverbond	heidendom	heidendom	volksregering	vloekverbond	souvereinen
landzaat	oproerigeid	verjaagde	landzaat	landzaat	volksregering
oproerigeid	handeldrijvend	handeldrijvend	oproerige	verjaagde	vrijheidlievend
jakobs	souvereinen	oproerigeid	vloekverbond	oproerige	opstand
isrels	vloekverwanten	landzaat	juda	vrijheidlievend	heidendom

Table 11
Nearest neighbors for the target word "volk" from the model with drift 10 (3/3)

1820	1830	1840	1850	1860	1870
oproerig	oproerig	oproerig	oproerig	oproerig	volks
volks	volks	volks	volks	volks	oproerig
oproer	oproer	koningdom	koningdom	koningdom	koningdom
koningdom	muitzucht	oproer	muitziek	muitziek	muitziek
muitzucht	koningdom	muitziek	muitzucht	muitzucht	muitzucht
muitziek	muitziek	muitzucht	oproer	landzaat	landzaat
israël	amalek	amalek	amalek	muitend	muitend
amalek	muitend	oproers	oproers	oproers	oproers
muitend	landzaat	muitend	muitend	oproer	oproer
oproers	oproers	landzaat	landzaat	amalek	vrijheidlievend
landzaat	israël	israël	sanhedrin	onafhanklijkheid	vloekverbond
sanhedrin	sanhedrin	sanhedrin	onafhanklijkheid	sanhedrin	onafhanklijkheid
vloekverbond	onafhanklijkheid	onafhanklijkheid	israël	vrijheidlievend	amalek
handeldrijvend	vloekverbond	heidendom	vloekverbond	vloekverbond	leuzen
vrijheidlievend	volkren	vloekverbond	vrijheidlievend	leuzen	heidendom
oproerigeid	heidendom	vrijheidlievend	heidendom	heidendom	sanhedrin
souvereinen	vrijheidlievend	leuzen	heidnen	handeldrijvend	handeldrijvend
onafhanklijkheid	heidnen	volkren	volkren	volkren	volkren
heidendom	juda	handeldrijvend	leuzen	israël	israël
volksregering	vorst	heidnen	juda	krijgsgeweld	oorlogskunst

Table 12
Nearest neighbors for the target word "natie" from the model with drift 10 (1/3)

1700	1710	1720	1730	1740	1750
handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende
naäpen	naäpen	naäpen	naäpen	naäpen	naäpen
nationaliteit	nationaliteit	nationaliteit	nationaliteit	nationaliteit	nationaliteit
naburen	naburen	universiteiten	universiteiten	universiteiten	universiteiten
gerechtshoven	gerechtshoven	naburen	naburen	naburen	naburen
universiteiten	universiteiten	buitenlanders	buitenlanders	buitenlanders	buitenlanders
buitenlanders	buitenlanders	gerechtshoven	gerechtshoven	gouvernementen	gouvernementen
volksregering	volksregering	volksregering	gouvernementen	gerechtshoven	volksregering
gouvernementen	gouvernementen	gouvernementen	volksregering	volksregering	gerechtshoven
rechtbanken	rechtbanken	rechtbanken	rechtbanken	rechtbanken	rechtbanken
republiek	republiek	republiek	republiek	republiek	republiek
taaleigen	taaleigen	taaleigen	taaleigen	bevolkingen	constitutie
bevolkingen	bevolkingen	bevolkingen	bevolkingen	taaleigen	bevolkingen
oppermachtige	oppermachtige	grondwetten	constitutie	constitutie	grondwetten
grondwetten	grondwetten	constitutie	grondwetten	grondwetten	taaleigen
verdienstelijkste	constitutie	oppermachtige	verdienstelijkste	verdienstelijkste	verdienstelijkste
constitutie	verdienstelijkste	verdienstelijkste	volksvertegenwoordiging	vrijheden	vrijheden
volksvertegenwoordiging	volksvertegenwoordiging	volksvertegenwoordiging	oppermachtige	volksvertegenwoordiging	democratie
onbeschaafdste	onbeschaafdste	onbeschaafdste	onbeschaafdste	onbeschaafdste	voortestaan
staatsinrichting	staatsinrichting	nederlanderen	nederlanderen	voortestaan	nederlanderen

Table 13
Nearest neighbors for the target word "natie" from the model with drift 10 (2/3)

1760	1770	1780	1790	1800	1810
handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende
naäpen	naäpen	naäpen	naäpen	naäpen	naäpen
nationaliteit	nationaliteit	nationaliteit	nationaliteit	nationaliteit	nationaliteit
universiteiten	naburen	naburen	naburen	naburen	naburen
naburen	universiteiten	universiteiten	gouvernementen	gouvernementen	gouvernementen
gouvernementen	gouvernementen	gouvernementen	volksregering	volksregering	volksregering
buitenlanders	buitenlanders	buitenlanders	universiteiten	universiteiten	universiteiten
volksregering	volksregering	volksregering	buitenlanders	buitenlanders	buitenlanders
gerechtshoven	rechtbanken	rechtbanken	rechtbanken	gerechtshoven	gerechtshoven
rechtbanken	gerechtshoven	gerechtshoven	gerechtshoven	republiek	republiek
republiek	republiek	republiek	republiek	rechtbanken	rechtbanken
constitutie	grondwetten	grondwetten	grondwetten	grondwetten	grondwetten
grondwetten	bevolkingen	bevolkingen	bevolkingen	bevolkingen	bevolkingen
bevolkingen	vrijheden	vrijheden	vrijheden	nabuuren	constitutie
vrijheden	constitutie	constitutie	nabuuren	constitutie	nabuuren
verdienstelijkste	verdienstelijkste	nabuuren	constitutie	vrijheden	vrijheden
taaleigen	taaleigen	verdienstelijkste	voortestaan	nederlanders	nederlanders
democratie	democratie	voortestaan	verdienstelijkste	voortestaan	voortestaan
voortestaan	voortestaan	democratie	nederlanders	verdienstelijkste	verdienstelijkste
nederlanderen	nabuuren	taaleigen	spruitende	democratie	nederlander

Table 14
Nearest neighbors for the target word "natie" from the model with drift 10 (3/3)

1820	1830	1840	1850	1860	1870
handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende
naäpen	naäpen	naäpen	naäpen	naäpen	naäpen
nationaliteit	nationaliteit	nationaliteit	nationaliteit	nationaliteit	nationaliteit
naburen	naburen	naburen	naburen	gerechtshoven	gerechtshoven
gouvernementen	volksregering	volksregering	universiteiten	volksregering	volksregering
volksregering	gouvernementen	universiteiten	volksregering	universiteiten	universiteiten
universiteiten	buitenlanders	buitenlanders	gerechtshoven	naburen	buitenlanders
buitenlanders	universiteiten	gouvernementen	buitenlanders	buitenlanders	gouvernementen
gerechtshoven	gerechtshoven	gerechtshoven	gouvernementen	gouvernementen	naburen
rechtbanken	bevolkingen	bevolkingen	oppermachtige	oppermachtige	oppermachtige
republiek	grondwetten	constitutie	taaleigen	taaleigen	taaleigen
constitutie	constitutie	grondwetten	constitutie	constitutie	keizerrijks
bevolkingen	republiek	republiek	bevolkingen	keizerrijks	bederver
grondwetten	rechtbanken	rechtbanken	grondwetten	voortestaan	constitutie
voortestaan	voortestaan	verdienstelijkste	keizerrijks	bederver	voortestaan
vrijheden	verdienstelijkste	voortestaan	verdienstelijkste	bevolkingen	onbeschaafdste
nabuuren	nederlanderen	oppermachtige	voortestaan	grondwetten	nakomeling
nederlanders	oppermachtige	taaleigen	rechtbanken	verdienstelijkste	nederlander
verdienstelijkste	nederlanders	keizerrijks	bederver	onbeschaafdste	bevolkingen
nederlanderen	vrijheden	onbeschaafdste	republiek	nederlander	grondwetten

B. Top 20 Nearest Neighbors for the model with precion of the random drift of 10

Table 15
Nearest neighbors for the target word "vaderland" from the model with drift 1 (1/3)

1700	1710	1720	1730	1740	1750
geboortegrond	geboortegrond	geboortegrond	geboortegrond	geboortegrond	geboortegrond
land	volksbestaan	volksbestaan	volksbestaan	volksbestaan	volksbestaan
volksbestaan	land	land	onafhanklijkheid	onafhanklijkheid	volksgeest
behulpzame	getuchtigd	getuchtigd	dierbaarst	bataven	bataven
getuchtigd	behulpzame	onafhankelijkheid	vrijgevochten	volksgeest	gemenebest
dierbaarst	onafhankelijkheid	onafhanklijkheid	roemrijk	land	onafhanklijkheid
allerdierbaarst	allerdierbaarst	dierbaarst	bataven	nederland	vrijgevochten
burgerschap	dierbaarst	roemrijkst	roemrijkst	vrijgevochten	roemrijkst
dierbaarste	roemrijkst	behulpzame	nederland	roemrijkst	nederland
onvergeetbare	onafhanklijkheid	burgerschap	burgertrouw	dierbaarst	roemrijk
vrijgevochten	dierbaarste	allerdierbaarst	heldenkroost	getuchtigd	vrijheidlievend
onafhankelijkheid	burgerschap	gemenebest	dwingelands	gemenebest	onafhankelijkheid
broederhand	gemenebest	vrijgevochten	allerdierbaarst	vrijheidlievend	geboorteland
onafhanklijkheid	vrijgevochten	volksgeest	gemenebest	onafhankelijkheid	dierbaarst
vaderlanders	onvergeetbare	volksgeluk	volksgeest	roemrijk	getuchtigd
scheldestad	volksgeluk	geboorteland	burgerdeugd	burgertrouw	burgertrouw
burgeres	geboorteland	dierbaarste	heldenmoed	heldenkroost	eendrachtsband
gewanhoopt	volksgeest	bataven	vrijheidlievend	eendrachtsband	nederlanderen
nederland	burgeres	burgertrouw	getuchtigd	burgerdeugd	behulpzame
roemrijkst	vaderlanders	onvergeetbare	volksgeluk	geboorteland	oorlogsroem
geboorteland	weldenkende	weldenkende	onafhankelijkheid	vrijheidszin	vrijheidszin

Table 16
Nearest neighbors for the target word "vaderland" from the model with drift 1 (2/3)

1760	1770	1780	1790	1800	1810
geboortegrond	geboortegrond	geboortegrond	geboortegrond	geboortegrond	geboortegrond
volksbestaan	volksbestaan	volksbestaan	volksbestaan	volksbestaan	volksbestaan
volksgeest	bataven	land	land	nederland	nederland
land	volksgeest	nederland	nederland	bataven	bataven
gemenebest	land	bataven	gemenebest	land	land
bataven	onafhanklijkheid	volksgeest	bataven	onafhanklijkheid	moedergrond
geboorteland	vrijgevochten	onafhanklijkheid	volksgeest	volksgeest	onafhanklijkheid
nederland	gemenebest	vrijgevochten	onafhanklijkheid	geboorteland	eendrachtsband
roemrijkst	roemrijk	eendrachtsband	vrijgevochten	eendrachtsband	vaderlande
roemrijk	roemrijkst	gemenebest	eendrachtsband	vrijgevochten	nederlanders
vrijgevochten	nederland	vrijheidlievend	geboorteland	moedergrond	vlaanderland
vrijheidlievend	vrijheidlievend	roemrijkst	vrijheidlievend	vaderlande	vrijgevochten
onafhanklijkheid	eendrachtsband	burgertrouw	roemrijkst	belgie	volksgeest
nederlanderen	vaderlande	roemrijk	vaderlande	gemenebest	burgerschap
nederlanders	geboorteland	moedergrond	belgie	roemrijkst	dierbaarst
onafhankelijkheid	burgertrouw	burgerdeugd	moedergrond	burgerschap	belgie
vaderlande	nederlanders	vaderlande	burgerschap	engelands	geboorteland
dierbaarst	dierbaarst	heldenkroost	engelands	nederlanders	roemrijkst
vaderlanders	oorlogsroem	oorlogsroem	onafhankelijkheid	vlaanderland	onbeklemde
oorlogsroem	burgerdeugd	nederlanders	dierbaarst	burgertrouw	burgertrouw

Table 17
Nearest neighbors for the target word "vaderland" from the model with drift 1 (3/3)

1820	1830	1840	1850	1860	1870
geboortegrond	geboortegrond	geboortegrond	geboortegrond	geboortegrond	geboortegrond
volksbestaan	bataven	nederland	bataven	bataven	vlaanderland
bataven	nederland	land	nederland	vlaanderland	bataven
moedergrond	volksbestaan	bataven	vlaanderland	land	eendrachtsband
onafhanklijkheid	land	onafhanklijkheid	onafhanklijkheid	nederland	vrijgevochten
land	moedergrond	volksbestaan	land	onafhanklijkheid	land
eendrachtsband	onafhanklijkheid	moedergrond	volksbestaan	eendrachtsband	neerland
nederland	dierbaarst	vlaanderland	moedergrond	moedergrond	moedergrond
vaderlande	eendrachtsband	eendrachtsband	eendrachtsband	neerland	onafhanklijkheid
vlaanderland	vlaanderland	dierbaarst	nederlanders	vrijgevochten	heldenkroost
dierbaarst	vaderlande	neerland	heldenkroost	volksbestaan	burgertrouw
onbeklemde	burgerschap	heldenkroost	vaderlande	heldenkroost	juiche
roemrijkst	heldenkroost	vaderlande	vrijgevochten	burgertrouw	nederland
nederlanders	vrijgevochten	burgerschap	neerland	nederlanders	onbeklemde
heldenkroost	roemrijkst	onbeklemde	dierbaarst	vaderlande	vaderlande
burgerschap	burgertrouw	burgertrouw	burgerschap	juiche	constand
vrijgevochten	volksgeest	vrijgevochten	burgertrouw	dierbaarst	dierbaarst
burgertrouw	nederlanders	nederlanders	juiche	onbeklemde	burgerschap
belgie	burgerdeugd	juiche	onbeklemde	burgerschap	bataaven
volksgeest	geboorteland	roemrijkst	roemrijkst	roemrijkst	vrijgevochte

Table 18
Nearest neighbors for the target word "volk" from the model with drift 1 (1/3)

1700	1710	1720	1730	1740	1750
israël	israël	israël	israël	israël	israël
isrel	oproerig	isrel	isrel	oproerig	oproerig
oproerig	isrel	oproerig	oproerig	isrel	isrel
oproer	oproer	heidnen	heidnen	oproer	oproer
heidnen	heidnen	oproers	oproer	sanhedrin	sanhedrin
egipten	oproers	isrels	sanhedrin	heidnen	heidnen
priesterdom	priesterdom	priesterdom	priesterdom	muitziek	oproers
sichem	muitziek	oproer	muitziek	priesterdom	muitziek
oproers	isrels	volkren	oproers	muitend	erfland
twistend	sanhedrin	muitziek	isrels	oproers	isrels
erfland	muitend	heidendom	muitend	isrels	muitend
egipte	twistend	sanhedrin	heidendom	erfland	egipten
heidendom	heidendom	erfland	erfland	heidendom	priesterdom
isrels	egipten	volken	volkren	twistend	kolk
sanhedrin	erfland	egipte	vloekverbond	volkren	twistend
josua	volkren	muitend	landzaat	landzaat	landzaat
volken	sichem	egipten	twistend	vloekverbond	volkren
volkren	egipte	landzaat	egipten	onroomsen	onroomsen
muitziek	landzaat	vloekverbond	amalek	amalek	handeldrijvend
muitend	vloekverbond	amalek	egipte	egipten	amalek
geboorteland	neerland	bataaven	burgerdeugd	landzaat	nassauws

Table 19
Nearest neighbors for the target word "volk" from the model with drift 1 (2/3)

1760	1770	1780	1790	1800	1810
israël	israël	oproerig	oproerig	oproerig	oproerig
oproerig	oproerig	israël	israël	sanhedrin	sanhedrin
isrel	isrel	sanhedrin	sanhedrin	israël	muitziek
sanhedrin	oproer	isrel	oproer	oproer	oproer
oproer	sanhedrin	oproer	muitziek	toegestroomde	oproers
oproers	oproers	muitziek	handeldrijvend	muitziek	handeldrijvend
heidnen	muitziek	oproers	oproers	oproers	landzaat
muitziek	landzaat	vorst	landzaat	handeldrijvend	isrel
kolk	heidnen	land	toegestroomde	isrel	muitend
priesterdom	volks	handeldrijvend	vorst	twistend	israël
isrels	onroomsen	isrels	isrel	volks	toegestroomde
landzaat	handeldrijvend	landzaat	volks	isrels	vrijheidlievend
egipten	land	muitend	twistend	landzaat	twistend
erfland	isrels	volks	priesterdom	onroomsen	isrels
vorst	vorst	heidnen	onroomsen	priesterdom	volks
muitend	twistend	toegestroomde	regeringloosheid	muitend	priesterdom
volkren	muitend	priesterdom	isrels	huurlingen	land
egipte	priesterdom	onroomsen	land	burgerrecht	onroomsen
twistend	erfland	erfland	erfland	land	regeringloosheid
volks	egipten	volkren	muitend	joodse	krijgsgevaar
geboorteland	neerland	bataaven	burgerdeugd	landzaat	nassauws

Table 20
Nearest neighbors for the target word "volk" from the model with drift 1 (3/3)

1820	1830	1840	1850	1860	187
muitziek	muitziek	sanhedrin	oproer	muitziek	muitziek
oproerig	sanhedrin	oproer	muitziek	oproer	volks
oproer	oproer	muitziek	volks	volks	sanhedrin
landzaat	landzaat	oproerig	oproerig	oproerig	oproer
sanhedrin	volken	priesterdom	sanhedrin	sanhedrin	vrijheidlievend
oproers	oproerig	volken	burgerrecht	vrijheidlievend	burgerrecht
volks	priesterdom	volks	oproers	burgerrecht	bastaardij
muitend	handeldrijvend	burgerrecht	volken	landzaat	natie
isrel	burgerrecht	vrijheidlievend	vrijheidlievend	bastaardij	vorst
handeldrijvend	isrel	landzaat	priesterdom	volken	tolk
onafhanklijkheid	volkren	handeldrijvend	landzaat	handeldrijvend	landzaat
vrijheidlievend	volks	regeringloosheid	bastaardij	isrel	onafhanklijkheid
muitzucht	vrijheidlievend	afgodsdienst	gewetensdwang	priesterdom	priesterdom
volken	oproers	natie	onafhanklijkheid	vorst	oproerig
priesterdom	onafhanklijkheid	bastaardij	isrel	gewetensdwang	isrel
krijgsgevaar	regeringloosheid	gewetensdwang	vorst	oproers	gewetensdwang
isrels	muitend	egipte	geestlijkheid	muitend	volksbestaan
vorst	vorst	land	natie	geestlijkheid	volken
volkren	tronen	heidendom	muitend	onafhanklijkheid	israël
toegestroomde	afgodsdienst	oproers	afgodsdienst	israël	heidendom

Table 21
Nearest neighbors for the target word "natie" from the model with drift 1 (1/3)

1700	1710	1720	1730	1740	1750
naburen	naburen	naburen	naburen	naburen	naburen
natiën	natiën	natiën	bevolkingen	bevolkingen	bevolkingen
bevolkingen	bevolkingen	bevolkingen	natiën	nabuuren	nabuuren
nabuuren	nabuuren	nabuuren	nabuuren	natiën	natiën
republiek	republiek	republiek	republiek	republiek	buitenlanders
buitenlanders	buitenlanders	buitenlanders	buitenlanders	buitenlanders	republiek
nederlanders	handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende	handeldrijvende
handeldrijvende	volksregering	volksregering	volksregering	volksregering	volksregering
aziatise	aziatise	aziatise	aziatise	aziatise	gerechtshoven
volksregering	nederlanders	gerechtshoven	nederlanders	gerechtshoven	aziatise
gerechtshoven	gerechtshoven	nationale	gerechtshoven	nederlanders	europeaanen
vrijheden	nationale	nederlanders	nationale	vrijheden	vrijheden
nationale	vrijheden	vrijheden	vrijheden	nationale	constitutie
gelukzoekers	gelukzoekers	gelukzoekers	gelukzoekers	europeaanen	nederlanders
nationaliteit	nationaliteit	nationaliteit	nationaliteit	gelukzoekers	nationaliteit
fabrikanten	koopsteden	koopsteden	koopsteden	nationaliteit	nationale
staatsmannen	staatsinrichting	staatsinrichting	europeaanen	naäpen	moedertaal
naäpen	naäpen	naäpen	naäpen	constitutie	naäpen
koopsteden	europeaanen	europeaanen	staatsinrichting	koopsteden	gelukzoekers
moedertaal	fabrikanten	fabrikanten	fabrikanten	moedertaal	chineezen

Table 22
Nearest neighbor for the target word "natie" from the model with drift 1 (2/3)

1760	1770	1780	1790	1800	1810
naburen	naburen	naburen	nederlanders	naburen	nabuuren
nabuuren	bevolkingen	bevolkingen	nabuuren	nabuuren	naburen
bevolkingen	nabuuren	nabuuren	nationale	nederlanders	bevolkingen
natiën	natiën	buitenlanders	naburen	bevolkingen	nederlanders
buitenlanders	republiek	natiën	voortbrengselen	natiën	natiën
republiek	nederlanders	nederlanders	landgenooten	buitenlanders	buitenlanders
nederlanders	buitenlanders	nationale	natiën	moedertaal	nationaliteit
europeaanen	volksregering	handeldrijvende	bevolkingen	nationaliteit	handeldrijvende
constitutie	handeldrijvende	gerechtshoven	moedertaal	nationale	gerechtshoven
gerechtshoven	gerechtshoven	vrijheden	buitenlanders	vrijheden	vrijheden
vrijheden	vrijheden	republiek	litteratuur	nederlanderen	moedertaal
volksregering	aziatise	aziatise	letterkunde	constitutie	europeaanen
moedertaal	europeaanen	nationaliteit	nationaliteit	handeldrijvende	constitutie
handeldrijvende	nationale	volksregering	nederlanderen	landgenooten	pericles
aziatise	constitutie	staatsbestuer	oorspronglijk	pericles	aziatise
voorouders	staatsbestuer	moedertaal	voortbrengsels	voortbrengselen	nederlanderen
chineezen	gelukzoekers	voortbrengselen	nederlander	aziatise	staatsmannen
naäpen	nationaliteit	constitutie	produkten	litteratuur	naäpen
grondwetten	naäpen	staatsmannen	vrijheden	gerechtshoven	volksregering
gelukzoekers	moedertaal	grondwetten	verbasteren	europeaanen	grondwetten

Table 23
Nearest neighbor for the target word "natie" from the model with drift 1 (3/3)

1820	1830	1840	1850	1860	1870
nabuuren	nabuuren	nabuuren	nabuuren	nabuuren	nederlanders
naburen	naburen	nederlanders	nederlanders	nederlanders	vrijheidlievend
bevolkingen	buitenlanders	naburen	constitutie	vrijheidlievend	nabuuren
buitenlanders	bevolkingen	buitenlanders	buitenlanders	constitutie	constitutie
nederlanders	natiën	constitutie	vrijheidlievend	buitenlanders	naburen
natiën	nederlanders	bevolkingen	gerechtshoven	naburen	gerechtshoven
europeaanen	handeldrijvende	handeldrijvende	handeldrijvende	gerechtshoven	nederlanderen
gerechtshoven	gerechtshoven	nederlanderen	naburen	nederlanderen	buitenlanders
handeldrijvende	nationaliteit	gerechtshoven	naäpen	aziatise	nationaliteit
vrijheden	europeaanen	natiën	vertegenwoordiging	handeldrijvende	naäpen
nationaliteit	nederlanderen	vrijheidlievend	aziatise	naäpen	grondwetten
nederlanderen	naäpen	nationaliteit	grondwetten	grondwetten	taaleigen
naäpen	aziatise	naäpen	nederlanderen	keizerrijks	volksvertegenwoordiging
pericles	pericles	pericles	pericles	pericles	voortestaan
aziatise	constitutie	aziatise	keizerrijks	nationaliteit	handeldrijvende
staatsmannen	vrijheidlievend	europeaanen	bevolkingen	bevolkingen	pericles
constitutie	vrijheden	grondwetten	nationaliteit	taaleigen	keizerrijks
volksregering	vertegenwoordiging	vertegenwoordiging	taaleigen	vertegenwoordiging	aziatise
landgenooten	staatsinrichting	keizerrijks	staatsinrichting	voortestaan	bevolkingen
grondwetten	keizerrijks	staatsinrichting	staatsmannen	staatsmannen	vertegenwoordiging

C. Absolute drift

Table 24
Top 25 of words with the largest absolute drift

Term	Drift
we	0.3864445388
hoolen	0.3762480617
stralen	0.3733189702
kievit	0.3710829616
witt	0.3491608202
noch	0.3384461701
mijnheer	0.3265607357
mus	0.3213006556
deeze	0.3206266463
zoals	0.3060097396
sijn	0.30450508
cato	0.3009181917
elizabeth	0.3000869751
zeiden	0.2948405743
ramen	0.2943577766
lens	0.2862321734
koenraad	0.2857837379
guller	0.2814703286
dezelve	0.2808713913
ten	0.27489236

D. Drift of target words with a precision drift of 1

Figure 8
The drift of target words and neutral words, in comparison to their position in the previous decade, from the model with a precision drift of 1