Deconstructing destruction: A Cognitive Linguistics perspective on a computational analysis of diachronic change

Karlien Franco
KU Leuven & FWO Flanders
karlien.franco@kuleuven.be

Kris Heylen
KU Leuven & Dutch Language Institute (INT)
kris.heylen@ivdnt.org

Mariana Montes
KU Leuven
mariana.montes@kuleuven.be

Abstract

In this paper, we aim to introduce a Cognitive Linguistics perspective into a computational analysis of near-synonyms. We focus on a single set of Dutch near-synonyms, *vernielen* and *vernietigen*, roughly translated as 'to destroy', replicating the analysis from Geeraerts (1997) with distributional models. Our analysis, which tracks the meaning of both words in a corpus of 16th-20th century prose data, shows that both lexical items have undergone semantic change, led by differences in their prototypical semantic core.

1 Introduction

This paper aims to stimulate further convergence between Cognitive Linguistics approaches to language change and computational methods for semantic change. Cognitive Linguistics is a contemporary linguistic paradigm that assumes that linguistic knowledge is rooted in general cognitive capabilities, that language is shaped by usage and that meaning entails conceptualization (Dabrowska and Divjak, 2015). It pays particular attention to the interaction between semasiological change, whereby a word's meaning changes over time, and onomasiological change, whereby the semantic configuration of a set of (near-)synonyms is reorganized over time. In this paper we focus on a single set of Dutch near-synonyms, *vernielen* and *vernietigen*, roughly translated as 'to destroy'. We replicate the analysis of these verbs from Geeraerts (1997) with distributional models. The analysis, which tracks the meaning of both words in a corpus of 16th-20th century prose data, shows that both lexical items have undergone semantic change led by differences in their prototypical semantic core, as predicted by Geeraerts' work.

2 Semasiology and onomasiology in Cognitive Linguistics

Following from "The cognitive commitment", which entails that the description of human language should be congruent with what is known about cognition within and outside of linguistics (Lakoff, 1990), Cognitive Linguistics aims to be a psychologically plausible model. Language is, thus, primarily studied as a means to communicate – a way to convey and process meaning. Furthermore, a maximalist, non-reductionist perspective on linguistic knowledge is assumed. Language systems are considered to be "reflections of general conceptual organization, categorization principles, processing mechanisms, and experiential and environmental influences" (Geeraerts and Cuyckens, 2007, 3). The movement, therefore, places a large emphasis on meaning.

Geeraerts et al. (1994) examine the structure of lexical variation in the use of clothing terminology in Dutch. Crucially, it is the first study to systematically emphasize the importance of two distinctions. On the one hand, it shows that in order to obtain a full picture of the structure of lexical variation, semasiological research should be complemented with an onomasiological approach. The semasiological perspective examines the range of applications of a particular expression. Semasiology is, for this reason, often defined as research into the meaning of a particular item: given a particular word or expression, what are the referents to which the word applies? In the case of the Dutch word *monitor*, for instance, a semasiological analysis would reveal that it can refer both to a SUPERVISOR, and to a COMPUTER SCREEN (see Heylen et al., 2015). The onomasiological perspective investigates naming rather than meaning. An onomasiological approach, thus, starts from a particular

Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change, pages 23 - 32

(type of) referent or concept and determines which names exist or can be used to refer to the referent. For instance, an onomasiological analysis of the concept COMPUTER SCREEN in Dutch would reveal that both *monitor* and *computerscherm* can be used to express this concept.

On the other hand, Geeraerts et al. (1994) was the first study to make the importance of the interaction between four different types of lexical variation for the structure of the lexicon explicit. First, it examines semasiological variation, the situation where a single lexical item can refer to more than one referent. For example, the lexical item *pants* can both be used to refer to a TWO-LEGGED TYPE OF OUTER GARMENT (IN GENERAL), but also to a more specific referent, viz. MEN'S UNDERWEAR. The second and third types of lexical variation that are distinguished concern two varieties of onomasiological variation: conceptual onomasiological variation and formal onomasiological variation. Conceptual onomasiological variation concerns the situation where "a referent or type of referent may be named by means of various conceptually distinct lexical categories" (Geeraerts et al., 1994, 3-4). For example, to refer to a pair of BLUE JEANS, a language user can either choose to select a lexical item belonging to the concept BLUE JEANS and use a word like *jeans* or *blue jeans*, or (s)he can conceptualize the referent as a type of PANTS, a superordinate concept, and call the denotatum *trousers* or *pants*. Formal onomasiological variation occurs when a choice has to be made between different synonymous expressions for a referent. In the blue jeans example, this would involve determining the relative frequency of the terms *jeans* versus *blue jeans* versus *trousers* versus *pants*. Finally, it shows how contextual variation can be at play both at the semasiological and onomasiological level. Contextual variation (also called speaker and situation related variation) is broadly defined: it includes both the relatively stable lectal properties of the interlocutors involved (like their gender or their nationality), but also transient situation-related features, like the register of the speech event (Geeraerts et al., 2010, 8). For the (onomasiological) blue jeans example, for instance, contextual variation may take the form of determining whether older people are more likely to refer to the concept as *blue jeans*.

The program laid out in Geeraerts et al. (1994) was applied to diachronic change in Geeraerts (1997). An important finding of this work is that semasiological and onomasiological variation and change are not independent of each other: semasiological changes also affect the onomasiological structure of a language.

3 Destructive verbs in Dutch

A classical analysis of the verbs meaning 'to destroy' in Dutch is presented in (Geeraerts, 1997, 1985, 1988). In these papers, Geeraerts analyzes *vernielen* and *vernietigen* in 19th century data taken from the citations corpus for largest historical dictionary of Dutch, the *Woordenboek der Nederlandsche taal* 'Dictionary of the Dutch Language'. Etymologically, the near-synonyms do not have the same root. *Vernielen* is a verb formed with a verbalizing prefix *ver-*, and *niel*, an obsolete Dutch adjective that roughly translates to 'down to the ground'. The literal meaning of the verb *vernielen* is then 'to throw down to the ground, to tear down'. *Vernietigen*, in contrast, is based on a verbalizing prefix *ver-*, with the adjective *nietig* (which itself comes from *niet* 'not, nothing' + a suffix *-ig*). The meaning of *vernietigen* is then 'to annihilate, to bring to naught'.

Despite these divergent sources, Geeraerts shows that the near-synonyms can be used in similar contexts by the same author in the 19th century. For instance, in examples 1 and 2 (adapted from Geeraerts, 1988, 30-31), *vernielen* and *vernietigen* occur in the context of a material artefact (a part of a building) being destroyed. He discusses many more examples that clearly show that the verbs are interchangeable in 19th century Dutch.

1. Dat huis ... werd ... tot den grond toe **vernield** (Veegens, Hist. Stud. 2, 282, 1869). [This house was demolished down to the ground.]

2. De vrijheidsmannen [hebben] ... het wapen des stichters in den voorgevel met ruwe hand **vernietigd** (Veegens, Hist. Stud. 1, 125, 1864). [The freedom fighters demolished the founder's arms in the facade with their rough hands.]

Overall, three semantic groups of uses for *vernielen* and *vernietigen* can be distinguished in his data (Table 1): concrete uses, abstract uses and personal uses. In the group of personal uses there is also a special case where an army is destroyed. This use can be considered to hold a middle position

With regard to concrete things
To demolish parts of buildings
To destroy other human artefacts
To destroy natural objects
With regard to concrete things
To annihilate existing situations, characteristics etc.
To prevent the execution of plans, intentions, etc.
With regard to persons
To kill someone
To undermine someone's physical health
To undermine someone's psychological well-being
To defeat groups of armed men or armies

Table 1: Uses of *vernielen* and *vernietigen*. Adapted from Geeraerts (1997, 191-192)

between the abstract (collective army) and personal (an individual soldier) contexts.

While the verbs are found in similar contexts, a crucial point of Geeraerts' work is that the prototypical cores of the verbs differ. More specifically, *vernielen* prototypically occurs with concrete uses, such as destroying parts of buildings. *Vernietigen* prototypically occurs in abstract contexts such as the complete annihilation of existing situations or plans. Geeraerts also notes that, while both verbs can occur with instances of partial or complete destruction, *vernielen* is prototypically used in the partial destruction sense (e.g. when a building is destroyed by a fire, parts of the structure and ashes from the fire remain), whereas *vernietigen* often implies complete annihilation to naught (e.g. when a plan is destroyed, nothing remains). In addition, the difference between the concrete and abstract uses is also visible in the context of the 'destruction' of people: while *vernielen* occurs more with the more concrete sense of to kill someone, *vernietigen* is more often found in the more abstract contexts where someone's physical or mental health is affected.

In Montes et al. (2021), the near-synonyms were analyzed in synchronic contemporary newspaper data with distributional models. The analysis showed that, since the 19th century, the prototypical cores of *vernielen* and *vernietigen* have become even stronger and the verbs are no longer easily interchangeable in every context. A highly prototypical context for *vernielen* in the 21st century data is the destruction of (parts of) buildings by fire and *vernietigen* no longer occurs in this context. In contrast, for *vernietigen*, the cancellation of decisions or ideas by a governmental body makes up a large portion of the tokens in the corpus and *vernielen* is no longer possible there. Both variants seem to have retreated to their prototypical core. In the periphery of the semantics of the verbs, some new uses have come into existence (e.g. to destroy livestock, probably as a result of the industrialization of the food industry and the regulations installed by government to keep the industry safe for consumption).

The aim of this paper is to track the diachronic change in the nearly synonymous pair *vernielen* and *vernietigen* throughout time in Dutch using distributional models. Based on the results in Geeraerts (1997, 1985, 1988) for 19th century dictionary attestations, and Montes et al. (2021) for 21st century corpus data, we expect to find in our study of continuous diachronic corpus data from the 16th to 20th century, that the overlap or interchangeability between the verbs reduces over time and that the verbs will retreat to their prototypical cores more and more over time. This finding would confirm that semasiological and onomasiological change interact and that these types of changes can be retrieved automatically from diachronic corpus data. Methodologically, we investigate the usefulness of distributional models for diachronic changes in a pair of near-synonyms.

4 Data and methods

In the analysis, we use a corpus of prose texts from DBNL, the *Digitale Bibliotheek voor de Nederlandse Letteren* 'digital library for Dutch languages and literature'. Some information about the corpus can be found in Depuydt and Brugman (2019), though the corpus is not publicly available at this time. We specifically extracted all corpus texts tagged as prose in the metadata from the 16th, 17th, 18th, 19th and 20th century. Due to data sparseness, we combine the subcorpora for the 16th and 17th century in the analysis.

As no high-quality lemmatizers or PoS-taggers are as of yet available for historical Dutch, the only preprocessing we applied to the corpus was to transform the entire corpus to lower case and to automatically indicate sentence boundaries using the pretrained nltk sentence tokenizer (Bird and Loper, 2004).

Next, we extracted all tokens for *vernielen* and *vernietigen* (including inflected forms and spelling variants) from the four subcorpora and took a random sample of N = 400 tokens for each subcorpus.

Then, we constructed a single vector space model for the tokens in each subcorpus. The models that we built are based on the procedure outlined in Schütze (1998). More specifically, we construct a vector representation for each token in the corpus, using the words in the context of the tokens to construct the vectors (first order vectors). We supplement this information by also constructing a vector for each of the relevant context words (second order vectors). An association strength measure is used rather than raw co-occurrence frequencies, namely positive pointwise mutual information or PPMI (Church and Hanks, 1989; Bullinaria and Levy, 2007; Kiela and Clark, 2014). This procedure has the advantage that both the context words for a token, as well as their semantic similarity with other context words, is taken into account (see De Pascale, 2019 and Montes, 2021). This method represents an example of a context-counting distributional model, which we opt for here (rather than for a context-predicting method) because we want to keep the results as comparable as possible to the results obtained in Montes et al. (2021) on contemporary data. One line of future research is to replicate the results obtained here with diachronic contextualized word embeddings when they become available for Dutch.

The parameters that we used are largely based on the best model found in Montes et al. (2021) for the analysis of *vernielen* and *vernietigen* in 21st century newspaper data. However, since the diachronic corpus that we use is not lemmatized or PoS-tagged, the vectors represent word forms rather than lemmas and we do not use part-of-speech filters. Additionally, we decided to decrease the window size from 15 to 10 words to the left and right of the target token for the first-order context words because preliminary analyses revealed that in models with a broader window, too many irrelevant or noisy context features were included in the analysis (also due to the fact that PoS filters cannot be applied). The parameters settings that we used are the following:

- Bag-of-words model with a window size of 10 words to the left and right of each token.
- First-order context words: all wordforms [w+] with a frequency of at least 10 in the subcorpus. First-order contextwords are subsequently filtered by their PPMI value with the target token: only words with PPMI > 2 are considered.
- Second order context words: 5000 most frequent wordforms [w+] in the subcorpus, excluding the first 100 wordforms, as these are usually function words rather than content words and therefore do not contribute a lot of semantic information.

The models were constructed with the nephosem Python library (QLVL, 2021). The result of this procedure is a token-by-context matrix with 5000 dimensions, where the dimensions represent second order context vectors, i.e. the vectors of the context words around each token. Since for some tokens no relevant context words are found with our parameter settings, these tokens are excluded from the remainder of the analysis. For clustering and visualization, we transformed this matrix into a square distance matrix by computing the cosine vector of each pair of token-level vectors, without further dimensionality reduction beforehand. Thus, this final matrix describes the dissimilarity between the vector of each token and all other token vectors in the subcorpus. As a next step, we submitted each model to a clustering procedure in R (R Core Team, 2020). We used hierarchical clustering (Ward method), distinguishing four clusters, following the procedure in Montes et al. (2021) for maximal comparability.

Finally, we analyzed each cluster per subcorpus basing ourselves on a procedure outlined in Montes (2021) that is available in the Python library semasioFlow (Montes, 2022). The procedure consists of a number of steps. First, the relevant context words, on which the token vectors are based, are extracted from the model data. Then, after each token is assigned to a cluster, we calculate how often a specific context word occurs within a particular cluster and outside of the cluster. Using this information, we can calculate which context words have an exceptionally high frequency in each cluster and therefore represent the semantics of each cluster well. In the analysis, we will only consider context words for which at least 50% of their occurrences are within the cluster of interest.

5 Results

Figure 1 (see Appendix A) shows the visualisations of the models, with one panel per subcorpus. Plot symbols show the variants (*vernielen* versus *vernietigen*) and colours indicate the clusters. The figure shows that over time, *vernielen* and *vernietigen* are distinguished more clearly by the models.

While in the 16th/17th century, there is still quite some overlap between the variants, indicating that they are still interchangeable, in the 18th century *vernielen* mostly occurs at the bottom left of the plot and *vernietigen* at the top right. In the 19th century, *vernielen* is found in the left side of the plot and *vernietigen* mostly in the bottom right. By the 20th century, the variant *vernielen* had decreased dramatically in frequency and *vernietigen* takes up most of the figure. Only one cluster remains where *vernielen* is dominant: cluster 4 at the bottom right.

Tables 2-5 show an overview of the most important context words per period and per cluster, obtained with the procedure outlined above. Only context words with a frequency of more than 2 are shown, to avoid that infrequent words get too much weight in the interpretation. The first column also shows an interpretation of each cluster. The final columns indicate the relative and absolute proportion of each variant in the cluster.

In the first subcorpus (16th-17th century, Table 2), there are three clusters where *vernielen* clearly is the major variant (clusters 2, 3 and 4). It occurs in contexts related to killing persons, a small cluster with natural objects (no context words with frequency > 2) and concrete objects like ships and cities. The first and largest cluster (N = 182 tokens) is still quite diverse and both *vernielen* and *vernietigen* are possible. Thus, in the 16th and 17th century, *vernielen* and *vernietigen* are still mostly interchangeable, although there are already a few contexts where *vernielen* is preferred.

In the 18th century subcorpus (Table 3), the variants start receding to their prototypical core more. There are two clusters where *vernielen* is more frequent and two clusters where *vernietigen* takes over. Following the hypotheses outlined above, *vernielen* mostly occurs with concrete objects like buildings (cluster 4, consisting of tokens related to fires destroying parts of buildings). In addition, it seems to occur in passive tokens (with *werden* 'became, was') where persons are destroyed: cluster 2 contains some war-related lexemes like *vijand* 'enemy', *troepen* 'troups', *leger* 'army' and some lexemes related to people, such as *hunne* 'their' and *elkaar* 'each other'. In contrast, *vernietigen* occurs in tokens with abstract objects (cluster 3) and it is also the most frequent variant in the first cluster, which does not show a clear semantic picture. In most clusters, except for 4, both variants are still possible. The context words in cluster 4 show that *vernielen* has by now become the most preferred variant for the destruction of (parts of) buildings (often by fire).

In the 19th century (Table 4), which coincides with the data analyzed in Geeraerts (1997, 1985, 1988), there are three clusters where one variant takes over, but also one cluster where the variants are interchangeable. More specifically, *vernielen* remains the most frequent variant in contexts of the destruction of (parts of) buildings (by fire, cluster 1). In contrast with the 18th century subcorpus, *vernietigen* has by now taken over contexts related to the destruction of persons, including armies (cluster 2). In this cluster *zichzelf* 'hisself/herself/themselves' is the most frequent context word. This frequent use of the reflexive pronoun may indicate that the patient role for *vernietigen* in the 19th century is often the subject itself, or that it at least plays a major role. Finally, *vernietigen* also still occurs the most with abstract lexemes such as *vrijheid* 'freedom' (cluster 4). Cluster 3 only has one important context word, *waan* 'delusion', and both variants are possible in this cluster. The interpretation is not as clear as for the other clusters.

Finally, in the subcorpus for the 20th century (Table 5), *vernietigen* is much more frequent than *vernielen*. Only 108 tokens for *vernielen* occur in the complete 20th century subcorpus, but 446 are available for *vernietigen*. This may indicate that *vernielen* is on its way out, or that it is retreating to very specific contexts. The cluster analysis shows that there are still some clear contexts in which *vernietigen* is the preferred variant, but that in the 20th century data, not all clusters represent clear semantic differences. This may be partly related to the fact that *vernielen* has become very infrequent: most of the tokens that are modelled are for *vernietigen* and it is possible that the model distinguishes syntactic constructions rather than semantic contexts in which *vernietigen* can occur.

First, *vernietigen* is the most frequent variant in cluster 1, which is a diverse cluster, with the most frequent context word related to complete destruction (*geheel* 'completely'), but also consisting of other types of lexical items such as abstract concepts. In cluster 2, *vernietigen* is the most frequent variant as well. This is a semantic cluster with many war-related lexical items, although it also contains other concrete objects. The context words in cluster 3, where *vernietigen* is also the most

frequent variant, are mostly function words, such as adverbs and reflexive pronouns. This cluster is not determined by semantic similarity between the tokens, but rather by the type of construction the tokens occur in. The context words in cluster 4, where both *vernielen* and *vernietigen* are possible, are mostly related to (parts of) buildings. This is a clear change compared to the earlier data, where the destruction of parts of building correlated strongly with the use of *vernielen*. However, the context words in this cluster have quite a low frequency so likely not all tokens are related to the destruction of (parts of) buildings: perhaps *vernielen* has become so infrequent that even this prototypical use is not frequent enough anymore to be distinguished by the model and clustering procedure.

6 Discussion

The models for the four subcorpora show how the relationship between the near-synonyms *vernielen* and *vernietigen* has changed over time. Semasiologically, *vernielen* was the major variant in the 16th and 17th century, occurring in tokens related to the death of persons and concrete, natural objects. Over the course of the 18th century, it developed its prototypical meaning related to the destruction of (parts of) buildings, often by fire, and this meaning remained its core usage in the 19th century. By the 20th century, the verb had decreased in frequency and its prototypical core was no longer distinguishable from the data. *Vernietigen*, in contrast, was the less frequent variant in the 16th and 17th century and at that time, there were no clear contexts yet where the verb occurred. It was mostly found in a semantically diverse cluster where its near-synonym *vernielen* was possible as well. From the 18th century onwards, the verb started to increase in frequency and it developed its prototypical sense of being used with abstract objects. In the 19th century, it also started to invade contexts where *vernielen* was preferred before (specifically related to the death of persons and to war). In the 20th century data, we also found a syntactic cluster, consisting of function words that often occur in the context of *vernietigen*.

Onomasiologically, the analysis showcased how the nuances in the concept 'to destroy' evolve over time and have become more outspoken. For instance, the clusters related to the destruction of parts of buildings are not yet visible in the oldest data but they are important clusters in the more recent datasets. Similarly, the cluster with abstract objects is not yet distinguished by the analysis for the 16th and 17th century, but these objects form a cluster on their own in the 18th and 19th century data. Moreover, the analysis also showed how these particular nuances of meaning are typically expressed by a particular verb. In the visualization, for instance, there is clearly less overlap (or interchangeability) between the verbs in the later periods (except in the 20th century data, where *vernielen* is infrequent).

Thus, this case-study showcases an example of how formal onomasiological variation and conceptual onomasiological variation can interact. On the one hand, *vernielen* and *vernietigen* serve as formal alternatives in the largest cluster from the 16th and 17th century data. However, from the 18th century onwards, each verb increasingly retreats to its prototypical core. Arguably, they should therefore be considered conceptually distinct, prototypically referring to different nuances of meaning, even though they remain nearly synonymous.

Methodologically, our usage of distributional models combined with a cluster analysis and the method, developed in Montes (2021), to analyze the context words that are good representatives for the clusters, allowed us to show how both verbs changed semantically over time. The procedure employed was quite straightforward, using a single set of parameter settings to model tokens from four diachronic subcorpora. With this procedure, we extended the analyses in Montes et al. (2021) and Geeraerts (1997) to a much longer time span. Despite the fact that we used a completely different dataset (a continuous diachronic corpus rather than dictionary citations from the 19th century only) and analysis method (an automatic procedure rather than a manual linguistic analysis), the hypotheses outlined in Geeraerts (1997) were mostly confirmed. Further, this method allowed us to track semasiological change and to investigate how this interacts with onomasiological variation over time.

One shortcoming of the approach is that the ideal settings for the parameters need not be the same for other near-synonyms or for a comparable linguistic alternation in other languages. In fact, this is one of the major findings of Montes (2021), who showed that there is no direct link between a choice of parameters and the linguistic phenomena that are revealed by a model constructed with the method proposed by Schütze (1998). Therefore, while in

Table 2: 16th & 17th century

Cluster	Context words	Variants
1 (diverse)	7: alles 'everything', geheel 'completely'; 4: natuur 'nature', geluk 'luck', duizend 'thousand', veranderingen 'changes', zonder 'without', werden 'became (pl.)', schulden 'debts', werd 'became (sg.)'; 3: gramschap 'wrath', beeld 'statue, picture', vorsten 'monarchs', oogenblik 'moment', kunt 'can', word 'become (sg.)', plantagiën 'plantations', nieuwe 'new', compagnie 'company', dezelve 'itself'	vernielen: 0.44 (80), vernietigen: 0.56 (102)
2 (TO KILL PERSONS)	10: dese 'this'; 5: doot 'death'; 4: desen 'this', t 'it'; 3: wet 'law', sulcke 'this', selve 'self', vyanden 'enemies', Christi '(of) Christ', dooden 'to kill', omme 'in order to', macht 'power', verlaten 'to leave', sonde 'sin'	**vernielen**: 0.84 (69), vernietigen: 0.16 (13)
3 (NATURAL OBJECTS)	/	**vernielen**: 0.90 (18), vernietigen: 0.10 (2)
4 (CONCRETE OBJECTS)	5: schepen 'ships'; 4: vernielen 'to destroy', steden 'cities', vloot 'fleet'; 3: zwaert 'sword', bergen 'mountains'	**vernielen**: 0.83 (45), vernietigen: 0.17 (9)

Table 3: 18th century

Cluster	Context words	Variants
1 (diverse)	5: daardoor 'because of'; 3: worde 'become (pl.)', gansch 'completely', hoop 'hope'	vernielen: 0.31 (12), **vernietigen**: 0.69 (27)
2 (TO KILL PERSONS + WAR)	12: werden 'become (pl.)'; 4: hunne 'their'; 3: elkaâr 'each other', vijand 'enemy', troepen 'troups', leger 'army', gebroken 'broken', slag 'battle', vloot 'fleet', oogst 'harvest'	**vernielen**: 0.61 (60), vernietigen: 0.39 (39)
3 (ABSTRACT OBJECTS)	5: invloed 'influence'; 4: zedelijk 'virtuous', kracht 'strength', macht 'power', revolutie 'revolution', bestaan 'existence, to exist', vrijheid 'freedom'	vernielen: 0.20 (14), **vernietigen**: 0.80 (56)
4 ((PARTS OF) BUILDINGS (FIRE))	6: brand 'fire'; 4: stad 'city'; 3: kerken 'churches', huizen 'houses', steden 'cities'	**vernielen**: 0.97 (35), vernietigen: 0.03 (1)

Table 4: 19th century

Cluster	Context words	Variants
1 ((parts of) buildings (fire))	5: huis 'house'; 4: brand 'fire'; 3: grond 'ground', boel 'things', vlammen 'flames'	**vernielen**: 0.84 (38), vernietigen: 0.16 (7)
2 (TO KILL PERSONS + WAR)	6: zichzelf 'hisself/herself/themselves', volkomen 'completely', steden 'cities'; 5: werden 'became (pl.)', leger 'army', vloot 'fleet', schepen 'ships'; 3: zorgvuldig 'carefully', gedeeltelijk 'partly', volledig 'completely', willen 'to want', brieven 'letters'	vernielen: 0.36 (35), **vernietigen**: 0.64 (63)
3 (ABSTRACT OBJECTS?)	3: waan 'delusion'	vernielen: 0.56 (15), vernietigen: 0.44 (12)
4 (ABSTRACT OBJECTS)	53: vrijheid 'freedom'	vernielen: 0.08 (3), **vernietigen**: 0.92 (37)

Table 5: 20th century

Cluster	Context words	Variants
1 (diverse)	5: geheel 'completely'; 4: bestaan 'existence, to exist'; 3: daardoor 'because of', groepen 'groups', rede 'reason', zulke 'such', schoonheid 'beauty', natuur 'nature', waarde 'value', dreigt 'threatens'	vernielen: 0.18 (19), **vernietigen**: 0.82 (87)
2 (PERSONS + WAR)	8: oorlog 'war'; 5: werden 'became (pl.)', nadat 'after', hele 'whole', gehele 'whole', oplage 'edition'; 4: documenten 'documents', moesten 'had to', volk 'people'; 3: recht 'right', kaart 'map', zouden 'would', joodse 'jewish', steden 'cities', exemplaren 'samples', geworden 'become (participle)', goden 'gods', zestig 'sixty', wereldoorlog 'world war', europese 'european'	vernielen: 0.16 (19), **vernietigen**: 0.84 (97)
3 (FUNCTION WORDS)	14: alles 'everything'; 10: zelfs 'even'; 9: zichzelf 'hisself/herself/itself'; 7: niets 'nothing'; 6: uiteindelijk 'eventually'; 4: mens 'human'; 3: waarna 'after which', god 'god', erbij 'near it', erop 'on it', definitief 'definitive', jezelf 'yourself', onmogelijk 'impossible'	vernielen: 0.18 (19), **vernietigen**: 0.82 (88)
4 ((PARTS OF) BUILDINGS)	4: huis 'house', aarde 'earth'; 3: muren 'walls', stenen 'stones'	vernielen: 0.52 (17), vernietigen: 0.48 (16)

this contribution we focus on a single set of parameters settings that were shown to be useful in analyses of the same linguistic example in another century, an alternative approach, that has been successfully employed in Montes (2021), is to consider a broader number of parameter settings to analyze linguistic phenomena.

References

Steven Bird and Edward Loper. 2004. NLTK: The natural language toolkit. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 214–217, Barcelona, Spain. Association for Computational Linguistics.

John A. Bullinaria and Joseph P. Levy. 2007. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, 39(3):510–526.

Kenneth Ward Church and Patrick Hanks. 1989. Word association norms, mutual information, and lexicography. In *ACL '89: Proceedings of the 27th Annual Meeting on Association for Computational Linguistic*, pages 76–83.

Ewa Dabrowska and Dagmar Divjak. 2015. Introduction. In Ewa Dabrowska and Dagmar Divjak, editors, *Handbook of Cognitive Linguistics*. De Gruyter Mouton, Berlin.

Stefano De Pascale. 2019. *Token-based vector space models as semantic control in lexical sociolectometry*. Doctoral dissertation, Leuven: KU Leuven.

Katrien Depuydt and Hennie Brugman. 2019. Turning Digitised Material into a Diachronic Corpus: Metadata Challenges in the Nederlab Project. In *DATeCH2019*, pages 169–173, New York.

Dirk Geeraerts. 1985. Preponderantieverschillen bij bijnasynoniemen. *De Nieuwe Taalgids*, 78:18–27.

Dirk Geeraerts. 1988. Where does prototypicality come from? In Brygida Rudzka-Ostyn, editor, *Topics in Cognitive Linguistics*, pages 207–229. John Benjamins, Amsterdam/Philadelphia.

Dirk Geeraerts. 1997. *Diachronic prototype semantics: a contribution to historical lexicology*. Oxford University Press, Oxford.

Dirk Geeraerts and Hubert Cuyckens. 2007. Introducing Cognitive Linguistics. In Dirk Geeraerts and Hubert Cuyckens, editors, *The Oxford Handbook of Cognitive Linguistics*, pages 3–21. Oxford University Press, Oxford.

Dirk Geeraerts, Stefan Grondelaers, and Peter Bakema. 1994. *The Structure of Lexical Variation: Meaning, Naming, and Context*. De Gruyter Mouton, Berlin, New York.

Dirk Geeraerts, Gitte Kristiansen, and Yves Peirsman, editors. 2010. *Advances in cognitive sociolinguistics*. Mouton de Gruyter, New York, N.Y.

K. Heylen, T. Wielfaert, D. Speelman, and D. Geeraerts. 2015. Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua*, 157:153–172.

Douwe Kiela and Stephen Clark. 2014. A Systematic Study of Semantic Vector Space Model Parameters. In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and Their Compositionality*, pages 21–30, Gothenburg. ACL.

George Lakoff. 1990. The invariance hypothesis. *Cognitive Linguistics*, 1(1):39–74.

Mariana Montes. 2021. *Cloudspotting. Visual analytics for distributional semantics*. Doctoral dissertation, Leuven: KU Leuven.

Mariana Montes. 2022. Montesmariana/semasioFlow: semasioFlow 0.1.0. Zenodo.

Mariana Montes, Karlien Franco, and Kris Heylen. 2021. Indestructible insights. A case study in distributional prototype semantics. In Gitte Kristiansen, Karlien Franco, Stefano De Pascale, Laura Rosseel, and Weiwei Zhang, editors, *Cognitive Sociolinguistics Revisited*, pages 251–264. De Gruyter Mouton, Berlin/Boston.

QLVL. 2021. *nephosem. Python module for type- and token-level distributional models.*

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Hinrich Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.

A Appendix

Figure 1: Data in four subcorpora