

# Simple, Interpretable and Stable Method for Detecting Words with Usage Change across Corpora

Hila Gonen<sup>\*1</sup> Ganesh Jawahar<sup>\*2</sup> Djamé Seddah<sup>2</sup> Yoav Goldberg<sup>1,3</sup>

<sup>1</sup>Department of Computer Science, Bar-Ilan University

<sup>2</sup>Inria Paris

<sup>3</sup>Allen Institute for Artificial Intelligence

hilagonn@gmail.com ganeshjwhr@gmail.com

yoav.goldberg@gmail.com djame.seddah@inria.fr

## Abstract

The problem of comparing two bodies of text and searching for words that differ in their usage between them arises often in digital humanities and computational social science. This is commonly approached by training word embeddings on each corpus, aligning the vector spaces, and looking for words whose cosine distance in the aligned space is large. However, these methods often require extensive filtering of the vocabulary to perform well, and—as we show in this work—result in unstable, and hence less reliable, results. We propose an alternative approach that does not use vector space alignment, and instead considers the neighbors of each word. The method is simple, interpretable and stable. We demonstrate its effectiveness in 9 different setups, considering different corpus splitting criteria (age, gender and profession of tweet authors, time of tweet) and different languages (English, French and Hebrew).

## 1 Introduction

Analyzing differences in corpora from different sources (different time periods, populations, geographic regions, news outlets, etc) is a central use case in digital humanities and computational social science. A particular methodology is to identify individual words that are used differently in the different corpora. This includes words that have their meaning changed over time periods (Kim et al., 2014; Kulkarni et al., 2015; Hamilton et al., 2016b; Kutuzov et al., 2018; Tahmasebi et al., 2018), and words that are used differently by different populations (Azarbondy et al., 2017; Rudolph et al., 2017). It is thus desired to have an *automatic, robust* and *simple* method for detecting such potential changes in word usage and surfacing them for human analysis. In this work we present such a method.

<sup>\*</sup>Equal contribution.

A popular method for performing the task (§4) is to train word embeddings on each corpus and then to project one space to the other using a *vector-space alignment* algorithm. Then, distances between a word-form to itself in the aligned space are used as an estimation of word usage change (Hamilton et al., 2016b). We show that the common alignment-based approach is unstable, and hence less reliable for the usage change detection task (§3,§7). In addition, it is also sensitive to proper nouns and requires filtering them.

We propose a new and simple method for detecting usage change, that does not involve vector space alignment (§5). Instead of trying to align two *different* vector spaces, we propose to work directly in the *shared vocabulary space*: we take the neighbors of a word in a vector space to reflect its usage, and consider words that have drastically different neighbours in the spaces induced by the different corpora to be words subjected to usage change. The intuition behind this approach is that words that are used significantly differently across corpora are expected to have different contexts and thus to have only few neighboring words in common. In order to determine the extent of the usage change of a word, we simply consider its top-*k* neighbors in each of the two corpora, and compute the size of the intersection of the two lists. The smaller the intersection is, the bigger we expect the change to be. The words are ranked accordingly.

The advantages of our method are the following:

1. **Simplicity**: the method is extremely simple to implement and apply, with no need for space alignment, hyperparameter tuning, and vocabulary filtering, except for simple frequency cutoffs.
2. **Stability**: Our method is stable, outputting similar results across different word embed-

dings trained on the same corpora, in contrast to the alignment-based approach.

3. **Interpretability:** The ranking produced by our method is very intuitive to analyze. Looking at the neighborhood of a word in the two corpora reveals both the meaning of the word in each, and the extent to which the word has changed.
4. **Locality:** The interpretability aspect is closely linked to the *locality* of the decision. In our approach, the score of each word is determined only by its own neighbours in each of the spaces. In contrast, in the projection based method the similarity of a pair of words after the projection depends on the projection process, which implicitly takes into account *all the other words in both spaces and their relations to each other*, as well as the projection lexicon itself, and the projection algorithm. This makes the algorithmic predictions of the projection-based methods opaque and practically impossible to reason about.

We demonstrate the applicability and robustness of the proposed method (§7) by performing a series of experiments in which we use it to identify word usage changes in a variety of corpus pairs, reflecting different data division criteria. We also demonstrate the cross-linguistic applicability of the method by successfully applying it to two additional languages beyond English: French (a Romance language) and Hebrew (a Semitic language).

We argue that future work on detecting word change should use our method as an alternative to the now dominant projection-based method. To this end, we provide a toolkit for detecting and visualizing word usage change across corpora.<sup>1</sup>

## 2 Task Definition

Our aim is to analyze differences between corpora by detecting words that are used differently across them. This task is often referred to as “detecting meaning change” (Azarboyad et al., 2017; Del Tredici et al., 2019).

However, we find the name “meaning change” to be misleading. Words may have several meanings in the different corpora, but different dominant sense in each corpus, indicating different use of the

---

<sup>1</sup>[https://github.com/gonenhila/usage\\_change](https://github.com/gonenhila/usage_change)

word. For this reason, we refer to this task as “detecting usage change”.

We define our task as follows: given two corpora with substantial overlapping vocabularies, identify words that their predominant use is different in the two corpora. The algorithm should return a ranked list of words, from the candidate that is most likely to have undergone usage-change, to the least likely.

Since the primary use of such algorithm is corpus-based research, we expect a human to manually verify the results. To this end, while the method does not need to be completely accurate, it is desirable that most of the top returned words are indeed those that underwent change, and it is also desirable to provide explanations or interpretations as to the usage of the word in each corpus. Lastly, as humans are susceptible to be convinced by algorithms, we prefer algorithms that reflect real trends in the data and not accidental changes in environmental conditions.

## 3 Stability

A desired property of an analysis method is *stability*: when applied several times with slightly different conditions, we expect the method to return the same, or very similar, results. Insignificant changes in the initial conditions should result in insignificant changes in the output. This increases the likelihood that the uncovered effects are real and not just artifacts of the initial conditions.

Recent works question the stability of word embedding algorithms, demonstrating that different training runs produce different results, especially with small underlying datasets. Antoniak and Mimno (2018) focuses on the cosine-similarity between words in the learned embedding space, showing large variability upon minor manipulations on the corpus. Wendlandt et al. (2018) make a similar argument, showing that word embeddings are unstable by looking at the 10-nearest neighbors (NN) of a word across the different embeddings, and showing that larger lists of nearest neighbors are generally more stable.

In this work, we are concerned with the stability of usage-change detection algorithms, and present a metric for measuring this stability. A usage-change detection algorithm takes as input two corpora, and returns a ranked list  $r$  of candidate words, sorted from the most likely to have changed to the least likely. For a stable algorithm, we expect different runs to return similar lists. While we do not

care about the exact position of a word within a list, we do care about the composition of words at the top of the list. We thus propose a measure we call **intersection@k**, measuring the percentage of shared words in the the top-k predictions of both outputs:

$$\text{intersection@k}(r_1, r_2) = \frac{|r_1^k \cap r_2^k|}{k} \quad (1)$$

where  $r_1$  and  $r_2$  are the two ranked lists, and  $r_i^k$  is the set of top  $k$  ranked words in ranking  $r_i$ .

A value of 0 in this measure means that there are no words in the intersection, which indicates high level of variability in the results, while a value of 1 means that all the words are in the intersection, indicating that the results are fully consistent. We expect to see higher **intersection@k** as  $k$  grows. This expectation is confirmed by our experiments in Section 7.2.

We measure the stability of the usage-change detection algorithms with respect to a change in the underlying word embeddings: we apply the **intersection@k** metric to two runs of the usage-change detection algorithm on the same corpus-pair, where each run is based on a different run of the underlying word embedding algorithm.

## 4 The Predominant Approach

The most prominent method for detecting usage change is that of [Hamilton et al. \(2016b\)](#), originally applied to detect shifts in dominant word senses across time. It is still the predominant approach in practice,<sup>2</sup> with recent works building upon it ([Yao et al., 2018](#); [Rudolph and Blei, 2018](#)). This method was also shown to be the best performing one among several others ([Schlechtweg et al., 2019](#)).

It works by training word embeddings on the two corpora, aligning the spaces, and then ranking the words by the cosine-distance between their representations in the two spaces, where large distance is expected to indicate significant change in meaning. We refer to this method as **AlignCos**.

The alignment is performed by finding an orthogonal linear transformation  $Q$  that, when given matrices  $X$  and  $Y$ , projects  $X$  to  $Y$  while minimizing the squared loss:

$$Q = \arg \min_Q \|QX - Y\|_2, \quad \text{s.t. } Q \text{ is orthogonal}$$

<sup>2</sup>This is also indicated by the large number of citations: 350 according to Google Scholar.

The rows of  $X$  correspond to embeddings of words in space A, while the rows of  $Y$  are the corresponding embeddings in space B. This optimization is solved using the Orthogonal Procrustes (OP) method ([Schönemann, 1966](#)), that provides a closed form solution.

Vector space alignment methods are extensively studied also outside of the area of detecting word change, primarily for aligning embedding spaces across language pairs ([Xing et al., 2015](#); [Artetxe et al., 2018b](#); [Lample et al., 2018a](#); [Artetxe et al., 2018a](#)). Also there, the Orthogonal Procrustes method is taken to be a top contender ([Lample et al., 2018b](#); [Kementchedjheva et al., 2018](#)).

### 4.1 Shortcomings of the alignment approach

**Self-contradicting objective.** Note that the optimization procedure in the (linear) alignment stage attempts to project each word to itself. This includes words that changed usage, and which therefore should not be near each other in the space. While one may hope that other words and the linearity constraints will intervene, the method may succeed, by mistake, to project words that did change usage next to each other, at the expense of projecting words that did not change usage further apart than they should be. This is an inherent problem with any alignment based method that attempts to project the entire vocabulary onto itself.

**Requires non-trivial filtering to work well.** In addition, the alignment-based method requires non-trivial vocabulary filtering to work well. For example, [Hamilton et al. \(2016b\)](#) extensively filter proper nouns. Indeed, without such filtering, proper-nouns dominate the top of the changed words list. This does not indicate real word usage change, but is an artifact of names being hard to map across embedding spaces. In that respect, it makes sense to filter proper nouns. However, *some* cases of word usage change do involve names. For example, the word “Harlem”, which is used as either a name of a neighborhood in NY or as a name of a basketball team, was detected by our method as a word whose usage changed between tweets of celebrities with different occupations (§7.1).

**Not stable across runs.** As we discuss in Section 3 and show in Section 7.2, the approach is not very stable with respect to different random seeds in the embeddings algorithm.

	Age		Gender		Occupation			Day-of-week		Hebrew		French	
	young	older	male	female	creator	sports	performer	weekday	weekend	2014	2018	2014	2018
#words	58M	116M	293M	126M	87M	132M	126M	142M	114M	42M	155M	867M	1B
#tweets	5M	8M	23M	9M	6M	11M	10M	9M	7M	4M	13M	82M	104M
#vocab	42K	73K	114K	69K	63K	66K	69K	81K	72K	84K	187K	263K	350K

Table 1: Statistics of the different splits.

## 5 Nearest Neighbors as a Proxy for Meaning

Rather than attempting to project two embedding spaces into a shared space (which may not even map 1:1), we propose to work at the shared vocabulary space. The underlying intuition is that words whose usage changed are likely to be interchangeable with different sets of words, and thus to have different neighbors in the two embedding spaces. This gives rise to a simple and effective algorithm: we represent each word in a corpus as the set of its top  $k$  nearest neighbors (NN). We then compute the score for word usage change across corpora by considering the size of the intersection of the two sets (not to be confused with `intersection@k` defined in Section 3):

$$\text{score}^k(w) = -|NN_1^k(w) \cap NN_2^k(w)| \quad (2)$$

where  $NN_i^k(w)$  is the set of  $k$ -nearest neighbors of word  $w$  in space  $i$ . Words with a smaller intersection are ranked higher as their meaning-change potential.

We only consider the words in the intersection of both vocabularies, as words that are rare in one of the corpora are easy to spot using the frequency in the two spaces, and do not neatly fit the definition of **usage** change.

Note that our method does not require extensive filtering of words – we only filter words based on their frequency in the corpus<sup>3</sup>.

We use a large value of  $k = 1000^4$  in practice, because large neighbor sets are more stable than small ones (Wendlandt et al., 2018), leading to improved stability for our algorithm as well.

**Limitations** Similar to previous methods, our method assumes high quality embeddings, and

<sup>3</sup>For English experiments we also filter stopwords according to the predefined list from NLTK.

<sup>4</sup>While this value may seem arbitrary, we tested several values in that range which yielded very similar results. However, the appropriate range may change when used with smaller corpora, or substantially different vocabulary sizes. We consider  $k$  to be the only hyperparameter of our method, and note that it is rather easy to set.

hence also a relatively large corpus. Indeed, in many cases we can expect large quantities of data to be available to the user, especially when considering the fact that the data needed is raw text rather than labeled text. Using a limited amount of data results in lower quality embeddings, but also with smaller vocabulary size, which might affect our method. For high-quality embeddings with small vocabulary sizes, we believe that changing  $k$  accordingly should suffice. Naturally, results will likely degrade as embeddings quality deteriorate.

It is also important to note that, like previous approaches, our method does not attempt to provide any guarantees that the detected words have indeed undergone usage change. It is only intended to propose and highlight **candidates** for such words. These candidates are meant to later be verified by a user who needs to interpret the results in light of their hypothesis and familiarity with the domain. Unlike previous methods, as we discuss in Section 7.4, our method also provides intuitive means to aid in such an interpretation process.

## 6 Experimental Setup

We compare our proposed method (NN) to the method of Hamilton et al. (2016b) described in Section 4 (**AlignCos**), in which the vector spaces are first aligned using the OP algorithm, and then words are ranked according to the cosine-distance between the word representation in the two spaces.<sup>5</sup> This method was shown to outperform all others that were compared to it by Schlechtweg et al. (2019).

We demonstrate our approach by using it to detect change in word usage in different scenarios. We use the following corpora, whose statistics are listed in Table 1.

We consider three demographics-based distinctions (age, gender, occupation), a day-of-week

<sup>5</sup>Some extensions may yield improved results (filtering out proper names, as done in Hamilton et al. (2016b), or jointly learning and aligning the spaces (Bamler and Mandt, 2017; Rudolph et al., 2017; Rudolph and Blei, 2018; Yao et al., 2018), but we stick to this setting as it is the most general out of this line of work, and the one most commonly used in practice, for which an open implementation is available.



based distinction, and short-term (4y) diachronic distinctions. We also compare to the longer-term (90y) diachronic setup of Hamilton et al. (2016b), which is based on Google books.

**Author Demographics** The Celebrity Profiling corpus (Wiegmann et al., 2019) consists of tweets from celebrities along with their traits such as age, gender and occupation. Based on these labels, we create the following splits: (1) **Age**: Young (birthyear 1990–2009) vs. Older (birthyear 1950–1969); (2) **Gender**: Male vs. Female; (3) **Occupation**: pairwise splits with Performer, Sports and Creator.

**Day-of-week** Yang and Leskovec (2011) collect 580 million tweets in English from June 2009 to February 2010, along with their time-stamps. As this is a fairly large corpus, we consider the tweets of a single month (November 2009). We create a split based on the Day-of-Week: weekday (tweets created on Tuesday and Wednesday) vs. weekend (tweets created on Saturday and Sunday). We remove duplicated tweets, as preliminary experiments revealed odd behavior of the representations due to heavily duplicated spam tweets.

**French Diachronic (4y, tweets)** Abitbol et al. (2018) compile a collection of tweets in French between the years 2014 and 2018. The authors utilize several heuristics based on the users’ spatial information to consider tweets from users based in French territory only. We use the 2014 and 2018 portions of the data, and create a split accordingly.

**Hebrew Diachronic (4y, tweets)** The Hebrew data we use is taken from a collection of Hebrew tweets we collected for several consecutive years, up to 2018. The collection was performed by using the streaming API and filtering for tweets containing at least one of the top 400 most frequent Hebrew words. We use the 2014 and 2018 portions of the data, and create a split accordingly.

**English Diachronic (90y, books)** For diachronic study on English corpora, we make use of the embeddings trained on Fiction from Google Books (Davies, 2015) provided by the authors of Hamilton et al. (2016b), specifically for the two years, 1900 and 1990. These embeddings are originally aligned using Orthogonal Procrustes and the words whose relative frequencies are above  $10^{-5}$  in both the time periods are ranked using cosine distance.

## 6.1 Implementation details

**Tokenization and Word Embeddings** We use 300 dimensions word2vec vectors with 4 words context window. Further details of embeddings algorithm and tokenization are available in the appendix.

**Vocabulary and Filtering** We perform frequency-based filtering of the vocabulary, removing stop words (the most frequent 200 words for each corpus, as well as English stop words as defined in `nlTK`<sup>6</sup>), as well as low frequency words (we discard the 20% least frequent words in each corpus, and require a minimum of 200 occurrences).

Notably, **we do not perform any other form of filtering**, and keep proper-nouns and person-names intact.

We consider neighbors having a raw frequency greater than 100 and identify 1000 such nearest neighbors ( $k = 1000$ ) to perform the intersection.

## 7 Results

### 7.1 Qualitative Evaluation: Detected Words

We run our proposed method and AlignCos (Hamilton et al., 2016b) on the different scenarios described in Section 6, and manually inspect the results. While somewhat subjective, we believe that the consistent success on a broad setting, much larger than explored in any earlier work, is convincing. We provide examples for two of the setups (English Diachronic and Performer vs. Sports), with the rest of the setups in the appendix. For each one, we list a few interesting words detected by the method, accompanied by a brief explanation (according to the neighbors in each corpus).

In addition, we depict the top-10 words our method yields for the **Age** split (Table 2), accompanied by the nearest neighbors in each corpus (excluding words in the intersection), to better understand the context. For comparison, we also mention the top-10 words according to the AlignCos method. Similar tables for the other splits are provided in the Appendix.

**Across all splits**, our method is able to detect high quality words as words that undergo usage change, most of them easily explained by their neighboring words in the two corpora. As expected, we see that the AlignCos method (Hamilton et al.,

<sup>6</sup><https://www.nltk.org/>



AGE (YOUNG VS. OLDER)	
NN	neighbors in each corpus
dem	dese, yuh, them, nuh, dey, ayye, dats, tha, betta, fuk repub, democrats, centrist, manchin, primaries, party’s, alp, dfl, gopers, repubs
dam	damm, mannnnn, mannnn, mane, huh, ahh, oo, buggin, koo, mannn dams, basin, river, dredging, reservoir, drainage, wastewater, sewerage, refinery, canal
rep	reppin, wear, allegiance, all-american, wildcat, alumni, tryout, hoosier, recruit, ua sen., congresswoman, chairwoman, co-chairs, gazelka, salazar, amb, comptroller, staffer, cong
assist	points, shutout, scoresheet, scored, pts, hatrick, sheet, nil, sacks,  assisting, contact, coordinate, locating, coordinating, administer, equip, consular, deploy, locate
pr	cameron, -pr, erik, lap, sargeant, laps, tundra, teamjev, caution, restart stunt, puerto, promotional, rico, creative, ploy, hire, spin, freelance, fema
fr	frfr, forreal, foreal, lmaooo, madd, tho, bck, bruhh, lmao, fwm pavone, incl, from, wrk, ger, joseph, covey, env, w, ans
joint	jawn, fusion, scorpion, sumn, spot, db, cb, joints, mgmt, fye high-level, convened, minsk, two-day, bilateral, counter-terrorism, delegations, asean, convene, liaison
mega	 , fantastic, simulator, macau, lotus, fuji, bmw, awesome, mclaren, fab gujarat, becos, multi-billion, gta, rupees, dollar, maharashtra, major, crores, multi-million
flow	beard, vibin, jeezy, drizzy, lite, mohawk, dreads, sauna, boomin, vibe illicit, influx, accumulation, moisture, absorb, overwhelm, heart’s, drains, curtail, diverting
icymi	superintendent, bureau, commissioner, spokesman, exec, state’s, prosecutor, reuters, montgomery, conway re-upping, reichert, newsmakers, sherrod, column, arizona’s, otl, holcomb, rundown, wrap-up
<b>AlignCos top-10</b>	<b>leo, whip, savage, nd, cole, pb, ace, carter, fr, bb</b>

Table 2: Top-10 detected words from our method (NN) vs. AlignCos method (last row), for corpus split according to the age of the tweet-author. Each word from our method is accompanied by its top-10 neighbors in each of the two corpora (Young vs. Older).

2016b) is highly sensitive to names, featuring many in the top-10 lists across the different splits. As opposed to AlignCos, our method is robust to global changes in the embedding space, since it looks at many neighbors. As a result, it is not sensitive to groups of words that “move together” in the embedding space (which might be the case with names).

**English (diachronic, 90y)** Top-100 words identified by our method cover all the words attested as real semantic shift in Hamilton et al. (2016b)’s top-10 except the word ‘wanting’. Specifically, three attested words, ‘gay’, ‘major’ and ‘check’ are present in our top-10, which also has more interesting words not present in Hamilton et al. (2016b)’s top-10 (1900 vs. 1990): **van** (captain vs. vehicle), **press** (printing vs. places), **oxford** (location vs. university). In addition, interesting words that came up in the top-30 list are the following: **headed** (body part vs. move in a direction), **mystery** (difficulty in understanding vs. book genre).

**Occupation (performer vs. sports)** Interesting words found at the top-10 list are the following: **cc** (carbon copy vs. country club), **duo** (duet vs. pair of people), **wing** (politics vs. football player position). In addition, interesting words that came up in the top-30 list are the following: **jazz** (music

genre vs. basketball team), **worlds** (general meaning vs. championships), **stages** (platforms vs. company(bikes)), **record** (music record vs. achievement), **harlem** (neighborhood vs. basketball team).

## 7.2 Quantitative Evaluation: Stability

We compare the stability of our method to that of the AlignCos method (Hamilton et al., 2016b) using the intersection@ $k$  metric, as defined in Section 3. We use  $k \in 10, 20, 50, 100, 200, 500, 1000$ .

In Figure 1(a) we plot the intersection@ $k$  for different values of  $k$  for all splits, with solid lines for the results of our method and dashed lines for the results of AlignCos method. It is clear that our method is significantly more stable, for all  $k$  values and across all splits. To better understand the parameters that affect the stability of the different methods, we also examine how the intersection changes with different values of frequency cut-off. In Figure 1(b) we plot intersection@100 as a function of the frequency cut-off (minimum word occurrences required for a word to be included in the ranking). Here, our method is again more stable for all corpus splits. In addition, our method is similarly stable, regardless the frequency cut-off, unlike the AlignCos method. We also examine how the size of NN lists considered for the intersection

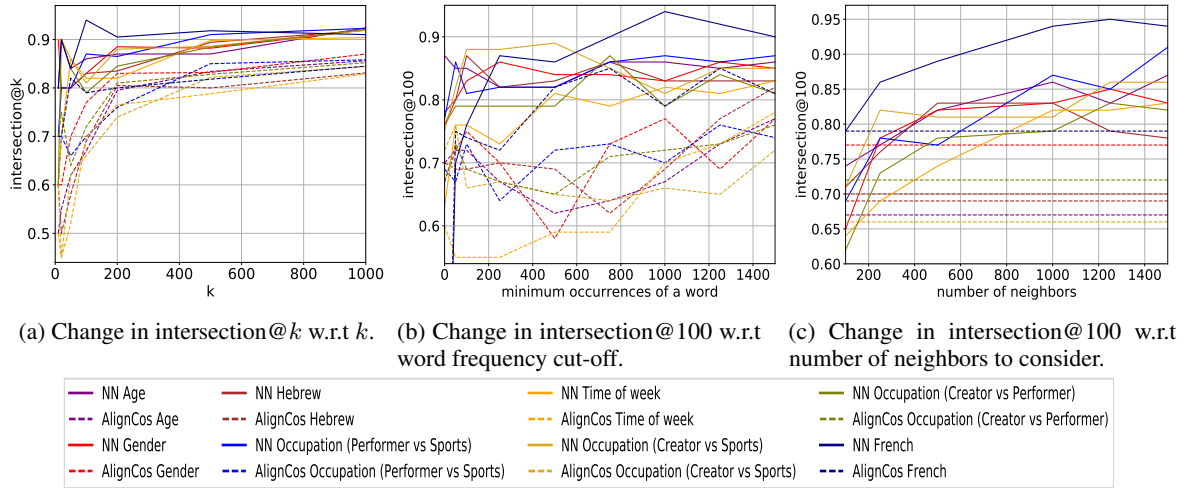


Figure 1: Stability plots. Solid lines: our method, dashed lines: AlignCos method.

affects the stability. In Figure 1(c) we plot the intersection@100 against number of neighbors taken into consideration using our method. We get that from around  $k = 250$ , our method is substantially more stable for all splits.

### 7.3 Quantitative Evaluation: DUREl and SUREl datasets

This field of semantic change suffers from lack of proper evaluation datasets, and there is no common benchmark that is being used. Two new datasets were recently introduced, and used to extensively compare between previous methods (Schlechtweg et al., 2019): the DUREl dataset (Schlechtweg et al., 2018) focuses on diachronic changes, while the SUREl dataset (Hätty et al., 2019) focuses on domain-based semantic changes. We use them to verify the quality of our results and compare against AlignCos (Hamilton et al., 2016b).

Both datasets include a limited number of German words, along with human annotations of the degrees of semantic relatedness between contexts of the words (across the different texts). However, they are not ideal as they are extremely limited (22 words each)<sup>7</sup>.

**Evaluation Metrics** Spearman correlation is the standard measure used in this field to compare between methods with respect to gold rankings. However, it is extremely important to note its limitations in this setting, since comparing to a very small gold ranking might be tricky. Specifically, it does

<sup>7</sup>For our experiments, we follow the setup of Schlechtweg et al. (2019) and use 19/21 words for DUREl/ SUREl respectively.

method	measure	SUREl	DUREl
AlignCos	spearman	0.800	0.814
NN	spearman	0.859	0.59
AlignCos	DCG	-4.5	-4.31
NN	DCG	-4.54	-4.3

Table 3: Results on DUREl and SUREl with NN and with AlignCos.

not take into account the global ranking of each method, but only the relative position of each of the gold words in each method’s ranking. For example, a method that ranks all the gold words at the bottom of the ranking (out of all the words in the vocabulary) in the same order, would be considered perfect, even though it is clearly not the case.

As a possible solution for this problem, we suggest to use Discounted Cumulative Gain (DCG), which better captures also global rankings. As opposed to Spearman, this measure takes into account not only the order of the words, but also their actual scores:

$$\text{DCG}(M) = \sum_{w \in W} \frac{\text{GoldScore}(w)}{\log_2(\text{rank}_M(w) + 1)} \quad (3)$$

where  $W$  are the words in the gold dataset, and  $M$  is the model being evaluated.

We report the results in Table 3. We compute AlignCos results with the best parameters reported in Schlechtweg et al. (2019)<sup>8</sup>. Our method outperforms AlignCos on SUREl, both when measur-

<sup>8</sup>We were unable to reproduce the exact results from the paper: spearman correlation of 0.866 and 0.851 on SUREl and DUREl, respectively.





word *clutch* in the **Gender** split, with cyan for female and violet for male, and the word *dam* in the **Age** split, with cyan for older and violet for young (in both cases they were no shared neighbours). We plot the projection of the words twice – one plot for each embedding space. We can see that, as expected, the neighboring words are distinct, and that the target word belongs to the respective neighborhood in each space. We conclude that this is a useful tool for interpreting the results of our model.

## 8 Related Work

Extensive work has been done on detecting word usage change across corpora that predated the alignment-based methods (Mitra et al., 2014; Jandt and Duh, 2014; Kenter et al., 2015; Ho et al., 2016; Frermann and Lapata, 2016).

In addition, two works are more closely related to our approach. In Azarbondy et al. (2017), the authors also use the neighbors of a word in order to determine its stability (and therefore, the extent to which it changes). Their best model combines the traditional alignment-based approach with weighting the neighbors according to their rank and their stability. The algorithm is iterative, and they update the stability of all the words in the vocabulary in each update step. Our method uses the neighbors of the words directly, does not include an iterative process, and does not rely on cosine-distance in the aligned embeddings. In addition, their method requires computation for the whole vocabulary, while other methods, including ours, usually allow querying for a single word.

Another work that considers the neighbors of the word in order to determine the extent of change is that of Hamilton et al. (2016a), in which they suggest a measure that is based on the changes of similarities between the target word and its neighbors in both spaces. They find that this method is more suitable for identifying changes that are due to cultural factors, rather than linguistic shift. This may serve as another motivation to move from the global measures to a local one.

Recent works (Giullianelli, 2019; Martinc et al., 2019) explored the possibility of modeling diachronic and usage change using contextualized embeddings extracted from now ubiquitous Bert representations (Devlin et al., 2019). Focusing on the financial domain, Montariol and Allauzen (2020) use, on top of Bert embeddings, a clustering method that does not need to predefine the number

of clusters and which leads to interesting results on that domain. Another approach from Hu et al. (2019) relies on the inclusion of example-based word sense inventories over time from the Oxford dictionary to a Bert model. Doing so provides an efficient fine-grained word sense representation and enables a seemingly accurate way to monitor word sense change over time. Most of those approaches could be easily used with our method, the inclusion of contextualized embeddings would be for example straightforward, we leave it for future work.

## 9 Conclusion

Detecting words that are used differently in different corpora is an important use-case in corpus-based research. We present a simple and effective method for this task, demonstrating its applicability in multiple different settings. We show that the method is considerably more stable than the popular alignment-based method popularized by Hamilton et al. (2016b), and requires less tuning and word filtering. We suggest researchers to adopt this method, and provide an accompanying software toolkit.

## Acknowledgments

We thank Marianna Apidianaki for her insightful comments on an earlier version of this work. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme, grant agreement No. 802774 (iEXTRACT), and from the the Israeli ministry of Science, Technology and Space through the Israeli-French Maimonide Cooperation programme. The second and third authors were partially funded by the French Research Agency projects ParSiTi (ANR-16-CE33-0021), SoSweet (ANR15-CE38-0011-01) and by the French Ministry of Industry and Ministry of Foreign Affairs via the PHC Maimonide France-Israel cooperation programme.

## References

Jacob Levy Abitbol, Márton Karsai, Jean-Philippe Magué, Jean-Pierre Chevrot, and Eric Fleury. 2018. Socioeconomic dependencies of linguistic patterns in twitter: A multivariate analysis. In *Proceedings of the 2018 World Wide Web Conference, WWW ’18*, pages 1125–1134.

- Maria Antoniak and David Mimno. 2018. [Evaluating the Stability of Embedding-based Word Similarities](#). *Transactions of the Association for Computational Linguistics*, 6:107–119.
- Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018a. [A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 789–798, Melbourne, Australia. Association for Computational Linguistics.
- Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. 2018b. [Unsupervised Neural Machine Translation](#). In *6th International Conference on Learning Representations, ICLR*.
- Hosein Azarbondy, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut, Maarten Marx, and Jaap Kamps. 2017. [Words Are Malleable: Computing Semantic Shifts in Political and Media Discourse](#). In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17*, pages 1509–1518.
- Robert Bamler 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, International Convention Centre, Sydney, Australia. PMLR.
- Mark Davies. 2015. [Corpus of Historical American English \(COHA\)](#).
- Marco Del Tredici, Raquel Fernández, and Gemma Boleda. 2019. [Short-Term Meaning Shift: A Distributional Exploration](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2069–2075, Minneapolis, Minnesota. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Lea Frermann and Mirella Lapata. 2016. [A Bayesian Model of Diachronic Meaning Change](#). *Transactions of the Association for Computational Linguistics*, 4(0).
- Mario Giullianelli. 2019. [Lexical semantic change analysis with contextualised word representations](#). Master's thesis, Institute for Logic, Language and Computation, University of Amsterdam, July.
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016a. [Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2116–2121, Austin, Texas. Association for Computational Linguistics.
- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016b. [Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.
- Anna Häddy, Dominik Schlechtweg, and Sabine Schulte im Walde. 2019. [SUREl: A Gold Standard for Incorporating Meaning Shifts into Term Extraction](#). In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (\*SEM 2019)*, pages 1–8, Minneapolis, Minnesota. Association for Computational Linguistics.
- Tin Kam Ho, Luis A. Lastras, and Oded Shmueli. 2016. [Concept Evolution Modeling Using Semantic Vectors](#). In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion*, pages 45–46. International World Wide Web Conferences Steering Committee.
- Renfen Hu, Shen Li, and Shichen Liang. 2019. [Diachronic sense modeling with deep contextualized word embeddings: An ecological view](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3899–3908, Florence, Italy. Association for Computational Linguistics.
- Adam Jatowt and Kevin Duh. 2014. [A Framework for Analyzing Semantic Change of Words Across Time](#). In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '14*, pages 229–238. IEEE Press.
- Yova Kementchedjheva, Sebastian Ruder, Ryan Cotterell, and Anders Søgaard. 2018. [Generalizing Procrustes Analysis for Better Bilingual Dictionary Induction](#). In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 211–220, Brussels, Belgium. Association for Computational Linguistics.
- Tom Kenter, Melvin Wevers, Pim Huijnen, and Maarten de Rijke. 2015. [Ad Hoc Monitoring of Vocabulary Shifts over Time](#). In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, CIKM '15*, pages 1191–1200. ACM.
- Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. [Temporal Analysis of Language through Neural Language Models](#). In

- Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 61–65, Baltimore, MD, USA. Association for Computational Linguistics.
- Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. [Statistically significant detection of linguistic change](#). In *Proceedings of the 24th International Conference on World Wide Web*, pages 625–635. International World Wide Web Conferences Steering Committee.
- Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. [Diachronic word embeddings and semantic shifts: a survey](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc’Aurelio Ranzato. 2018a. [Unsupervised machine translation using monolingual corpora only](#). In *Proceedings of ICLR*.
- Guillaume Lample, Alexis Conneau, Marc’Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018b. Word translation without parallel data. In *6th International Conference on Learning Representations, ICLR*.
- Laurens van der Maaten and Geoffrey Hinton. 2008. [Visualizing Data using t-SNE](#). *Journal of machine learning research*, 9(Nov):2579–2605.
- Matej Martinc, Petra Kralj Novak, and Senja Pollak. 2019. Leveraging contextual embeddings for detecting diachronic semantic shift. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv:1301.3781*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*.
- Sunny Mitra, Ritwik Mitra, Martin Riedl, Chris Biemann, Animesh Mukherjee, and Pawan Goyal. 2014. [That’s sick dude!: Automatic identification of word sense change across different timescales](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1020–1029, Baltimore, Maryland. Association for Computational Linguistics.
- Syrielle Montariol and Alexandre Allauzen. 2020. Étude des variations sémantiques à travers plusieurs dimensions. In *Actes de la conférence conjointe JEP-TALN 2020*, Nancy, France.
- Maja Rudolph and David Blei. 2018. [Dynamic Embeddings for Language Evolution](#). In *Proceedings of the 2018 World Wide Web Conference, WWW ’18*, pages 1003–1011.
- Maja Rudolph, Francisco Ruiz, Susan Athey, and David Blei. 2017. [Structured embedding models for grouped data](#). In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, pages 250–260, USA. Curran Associates Inc.
- Dominik Schlechtweg, Anna Hätty, Marco Del Tredici, and Sabine Schulte im Walde. 2019. [A Wind of Change: Detecting and Evaluating Lexical Semantic Change across Times and Domains](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 732–746, Florence, Italy. Association for Computational Linguistics.
- Dominik Schlechtweg, Sabine Schulte im Walde, and Stefanie Eckmann. 2018. [Diachronic Usage Relatedness \(DURel\): A Framework for the Annotation of Lexical Semantic Change](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 169–174, New Orleans, Louisiana. Association for Computational Linguistics.
- Peter H Schönemann. 1966. [A generalized solution of the orthogonal procrustes problem](#). *Psychometrika*, 31(1):1–10.
- Milan Straka and Jana Straková. 2017. [Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with udpipes](#). In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada. Association for Computational Linguistics.
- Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2018. [Survey of Computational Approaches to Lexical Semantic Change](#). *CoRR*, abs/1811.06278.
- Laura Wendlandt, Jonathan K. Kummerfeld, and Rada Mihalcea. 2018. [Factors Influencing the Surprising Instability of Word Embeddings](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2092–2102, New Orleans, Louisiana. Association for Computational Linguistics.
- Matti Wiegmann, Benno Stein, and Martin Potthast. 2019. [Celebrity Profiling](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2611–2618, Florence, Italy. Association for Computational Linguistics.
- Chao Xing, Dong Wang, Chao Liu, and Yiye Lin. 2015. [Normalized Word Embedding and Orthogonal Transform for Bilingual Word Translation](#). In *Proceedings of the 2015 Conference of the North*

*American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1006–1011, Denver, Colorado. Association for Computational Linguistics.

Jaewon Yang and Jure Leskovec. 2011. [Patterns of Temporal Variation in Online Media](#). In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 177–186, New York, NY, USA. ACM.

Zijun Yao, 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. ACM.



## A Implementation Details

**Tokenization** We tokenize the English, French and Hebrew tweets using ark-twokenize-py<sup>10</sup>, Moses tokenizer<sup>11</sup> and UDPipe (Straka and Straková, 2017), respectively. We lowercase all the tweets and remove hashtags, mentions, retweets and URLs. We replace all the occurrences of numbers with a special token. We discard all words that do not contain one of the following: (1) a character from the respective language; (2) one of these punctuations: “-”, “'”, “.”; (3) emoji.

**Word embeddings** We construct the word representations by using the continuous skip-gram negative sampling model from Word2vec (Mikolov et al., 2013a,b). We use the Gensim<sup>12</sup> implementation. For all our experiments, we set vector dimension to 300, window size to 4, and minimum number of occurrences of a word to 20. The rest of the hyperparameters are set to their default value.

For the **stability** experiments we run the embedding algorithm twice, each time with a different random seed.

## B Qualitative Evaluation: Detected Words

We show the top-10 words our method yields for each of the different splits, accompanied with the nearest neighbors in each corpus (excluding words in the intersection), to better understand the context. For comparison, we also show the top-10 words according to the AlignCos method. The splits are the following:

**English: 1900 vs. 1990** The list of top-10 detected words from our method (NN) vs. AlignCos method, for corpus split according to the year of the English text is displayed in Table 4.

**Age: Young vs. Older** The list of top-10 detected words from our method (NN) vs. AlignCos method, for corpus split according to the age of the tweet-author is displayed in Section 7. Interesting words found at the top-10 list are the following (young vs. older): **dem** (‘them’ vs. US political party), **dam** (‘damn’ vs. water barrier), **assist** (football contribution vs. help). In addition, interesting

<sup>10</sup><https://github.com/myleott/ark-twokenize-py>

<sup>11</sup>[https://www.nltk.org/\\_modules/nltk/tokenize/moses.html](https://www.nltk.org/_modules/nltk/tokenize/moses.html)

<sup>12</sup><https://radimrehurek.com/gensim/models/word2vec.html>

words that came up in the top-30 list are the following: **pc** (personal computer vs. Canadian party), **presents** (introduces vs. gifts), **wing** (general vs. political meaning), **prime** (general vs. political meaning), **lab** (school vs. professional).

**Gender: Male vs. Female** The list of top-10 detected words from our method (NN) vs. AlignCos method, for corpus split according to the gender of the tweet-author is displayed in Table 5. Interesting words found at the top-10 list are the following (male vs. female): **clutch** (grasping vs. female bag), **bra** (colloquial usage like ‘bro’ vs. female clothing), **gp** (grand prix event vs. general practitioner). In addition, interesting words that came up in the top-40 list are the following: **stat** (statistics vs. right away), **pit** (car-related vs. dog-related), **dash** (radio station vs. quantity), **pearl** (pearl harbor vs. gemstone and color).

**Occupation: Performer vs. Sports** The list of top-10 detected words from our method (NN) vs. AlignCos method, for corpus split according to the occupation (Performer vs. Sports) of the tweet-author is displayed in Table 6.

**Occupation: Creator vs. Sports** The list of top-10 detected words from our method (NN) vs. AlignCos method, for corpus split according to the occupation (Creator vs. Sports) of the tweet-author is displayed in Table 7. Interesting words found at the top-10 list are the following (creator vs. sports): **cc** (carbon copy vs. country club), **op** (event opening vs. operation), **wing** (politics vs. football player position), **worlds** (earth vs. world cup). In addition, interesting words that came up in the top-20 list are the following: **oval** (oval office vs. sports ground), **fantasy** (genre vs. fantasy football), **striking** (shocking vs. salient), **chilling** (frightening vs. relaxing), **fury** (book: fire and fury vs. British boxer).

**Occupation: Creator vs. Performer** The list of top-10 detected words from our method (NN) vs. AlignCos method, for corpus split according to the occupation (Creator vs. Performer) of the tweet-author is displayed in Table 8. Interesting words found at the top-10 list are the following (creator vs. performer): **dash** (travel vs. person), **presents** (introduces vs. gifts), **chapter** (book vs. movie). In addition, interesting words that came up in the top-30 list are the following: **cartoon** (cartoonist vs. movie), **scoop** (news story vs. ice cream),

**mega** (money vs. largeness), **sessions** (assembly vs. period).

**Time of week: Weekday vs. Weekend** The list of top-10 detected words from our method (NN) vs. AlignCos method, for corpus split according to the time of week (Weekday vs. Weekend) of the tweet is displayed in Table 9. Interesting words found at the top-10 list are the following (weekday vs. weekend): **cc** (credit card vs. carbon copy), **pitch** (presentation attribute vs. playing surface), **bond** (agreement vs. movie character). In addition, interesting words that came up in the top-30 list are the following: **sunday** (day of the week vs. vacation-related), **vp** (vice president vs. tv-series: True Jackson, VP), **third** (report-related vs. sports-related), **cliff** (first name vs. mountain cliff), **fight** (general meaning vs. boxing).

**French: 2014 vs. 2018** The list of top-10 detected words from our method (NN) vs. AlignCos method, for corpus split according to the year of the French text is displayed in Table 10. Interesting words found at the top-10 list are the following (2014 vs. 2018): **ia** (frequent misspelled contraction of “ya” in 2014, vernacular form of “il y a”, *there is*, vs. “intelligence artificielle”, *artificial intelligence*), **divergent** (the movie vs. the adjective). In addition, interesting words that came up in the top-30 list are the following: **pls** (contraction of the borrowing “please” vs. the acronym of “Position latérale de sécurité”, *lateral safety position*, which is now used as a figurative synonym for “having a stroke”). In the same vein, and tied to political debates, we note **apl** (contraction of “appel/appeler”, *call/to call* vs. controversial housing subsidies).

**Hebrew: 2014 vs. 2018** The list of top-10 detected words from our method (NN) vs. AlignCos method, for corpus split according to the year of the Hebrew text is displayed in Figure 4. Interesting words found at the top-10 list (2014 vs. 2018) are the following (we use transliteration accompanied with a literal translation to English): **beelohim-in god** (pledge word vs. religion-related) and **Kim-Kim** (First name vs. Kim Jong-un). In addition, interesting words that came up in the top-30 list are the following: **shtifat-washing** (plumbing vs. brainwashing), **miklat-shelter** (building vs. asylum (for refugees)), **borot-pit/ignorance** (plural of pit vs. ignorance).

English (1900 vs. 1990)	
NN	neighbors in each corpus
gay	cheery, humoured, apparel, natured, dresses, attire, neat, bright, genial, unusually lesbian, transgender, lesbians, katz, bisexual, bisexuals, coalition, gays, bi, gras
van	wyk, commented, sterne, skipper, south, simon, defarge, ned, island, carolina truck, helsing, luyden, luydens, pickup, toyota, jeep, porsche, volvo, der
press	pressed, publisher, papers, issues, dublin, circulation, thickest, wilson, paper, payment ams, belknap, harvester, wesleyan, newberry, westview, middletown, esp, harrington, gainesville
oxford	durham, albany, lincoln, sometime, ireland, john, canon, christ, bishops, newcastle clarendon, basingstoke, supervising, blackwell, 1921, researching, database, ibadan, walton, peruse
major	curtly, osborne, gordon, retorted, dryly, inspector, steele, chester, stewart, morris brigadier, factor, dramatist, producers, andre, schomburg, boswell, brian, biggest, insignia
2	vide, woodcuts, illustrations, peggy, demy, cloister, portrait, memoirs, baroness, allen rte, 767, tn, dresden, vols, 38225, bp, klingon, 1863, 98765432
cambridge	dublin, queens, glasgow, tutor, jesus, newcastle, christ, assistant, student, kent belknap, blackwell, 1921, persephone, harvester, hogarth, clarendon, ams, vols, esp
1	ornamental, woodcuts, dad, biography, section, demy, cent, 8vo, t, 3s xlibris, deduct, freepost, 345, 1001, 98765432, 350, 888, toulouse, bunkyo
new	revised, comer, institute, commonwealth, comers, development, insurance, illustrated, testament, magazine ungar, picayune, schocken, ams, crowell, atheneum, upstate, 10012, praeger, harrington
check	restrain, effort, balance, exertion, strove, readiness, restrained, gave, jerk, held cashier, update, checkbook, checks, payable, money, certificate, postal, brochure, lor
AlignCos Top-10	wanting, gay, check, starting, major, actually, touching, harry, headed, romance

Table 4: Top-10 detected words from our method (NN) vs. AlignCos method (last row), for corpus split according to the year of the text. Each word from our method is accompanied by its top-10 neighbors in each of the two corpora (1900 vs. 1990).

Gender (male vs. female)	
bra	bruuh, brah, bro, cuh, homie, boi, cuzzo, dawg, breh, brudda thong, jeans, strapless, leggings, tights, underwear, skirt, pants, sneakers, shorts
clutch	threes, walkoff, mookie, dingers, layups, midrange, game-winning, diaw, gwg, layup sequin, beaded, gown, dress, handbag, chiffon, headpiece, tote, sandal, swarovski
mm	cores, thickness, oled, diameter, deg, usb-c, ssd, dbo, gpu, cpu 🤔, huh, arizona, that's, errrr, bcz, thts, cc, 🤔, 🤔
mc	armand, dilla, rza, kuntryking, rapper, boney, riz, donald's, huss, dizzee obe, showstopper, groupie, fleming, thanks, hoff, cohost, honoree, harmon, reece
gp	motogp, thruxton, monza, indycar, dtm, snetterton, suzuka, hockenheim, criterium, wec physicians, pharmacists, clinical, procurement, ndis, insurers, nbn, tfl, hep, mh
keeper	midfielder, cech, krul, benteke, free-kick, freekick, aguero, defoe, benzema, goalscorer dynamo, goofball, hero, hustler, touche, stud, digger, nemesis, saver, ruler
nd	tht, iu, wvu, gtown, isu, wisco, ou, gng, huggs, byu minot, nh, ky, hoosier, farmers, heitkamp, ranchers, dakota, rural, ndans
hay	bales, doon, beech, hinton, blackwood, noches, ayer, mong, dartford, rooty beccy, goat, mclaren, portage, ale, glasto, grafton, daffodils, cornish, crap
steph	lebron, kyrie, klay, harden, draymond, rondo, melo, delly, dwade, korver chels, rach, leah, sam, liz, dani, trish, lovie, cait, kel
echo	homepod, orc, cortana, npc, oculus, undead, redstone, forked, emergent, echoed paradiso, avalon, asbury, hyde, sondeheim, colosseum, oasis, 🏹, empress, inconvenient
AlignCos Top-10	bra, mm, todd, bonnie, ralph, casey, stacey, gordon, lou, dana

Table 5: Top-10 detected words from our method (NN) vs. AlignCos method (last row), for corpus split according to the gender of the tweet-author. Each word from our method is accompanied by its top-10 neighbors in each of the two corpora (Male vs. Female).




Occupation (performer vs. sports)	
NN	neighbors in each corpus
blues	funk, reggae, b.b., boneshakers, folk, bluegrass, grooves, rhythm, trippers, moody hawks, leafs, rangers, sabres, bruins, tahs, fulham, knights, yotes, maroons
cc	 , lol sabathia, montclair, firestone, bethpage, isleworth, tourn, dorado, quail, riviera, westchester
dub	anime, subtitles, dubbing, dubbed, dlc, boxset, badman, rmx, miku, trax lakeshow, crunk, yessir, w, yessirr, ayeeee, win, ayeeeeee, yessirrr, yesir
bra	thong, panty, headband, panties, spanx, jeans, corset, uggs, tights, blouse bro, cuh, brodie, boi, dawg, brahh, breh, broo, cuzz, cuzo
track	rmx, tunes, album's,  , trax, single, sampler, instrumental, unreleased, song's racetrack, racin, field, slicks, velodrome, circuit, mtb, race, racing, sandown
wing	extremist, liberal, right-wing, fascists, leftist, conservative, propaganda, extremism, extremists, nationalists flank, footed, rear, wingers, fullback, retake, netting, seat, midfield, fullbacks
par	ghar, nahin, dekhna, mujhe, rahe, kiya, apne, naam, aaj, theek pars, bogey, birdie, holes, putts, hole, putted, fairway, birdied, sawgrass
mo	starlite, reeds, knuckleheads, bossier, rocke, kcmo, stafford, granada, hutchinson, rosemont bamba, tash,  , wesley, kev, mane, yessssssss, wes, yesssir, muzzy
ace	sweeeet, fantastic, amazeballs, rad, amaaaazing, exceptional, sweeeet, jez, amazing-, hoot sickkk, jb, robin, angel, stoner, ostrich, ayeeeeee, milly, homey, hustler
duo	supergroup, violinist, troupe, cardenas, stylings, cellist, baritone, multi-talented, vocalist, bassist tandem, northgate, dominant, keanu, hooker, wingers, rebounder, squads, superstar, jada
AlignCos Top-10	spencer, reed, dub, kurt, jerry, kirk, nova, watson, wa, curtis

Table 6: Top-10 detected words from our method (NN) vs. AlignCos method (last row), for corpus split according to the occupation of the tweet-author. Each word from our method is accompanied by its top-10 neighbors in each of the two corpora (performer vs. sports).

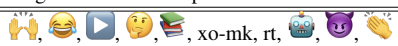
Occupation (creator vs. sports)	
NN	neighbors in each corpus
cc	 , montclair, firestone, bethpage, isleworth, tourn, dorado, quail, riviera, westchester, vero
op	nel, reeva, roux, hoare, pathologist, shauna, baden-clay, ed, nedrow, barrister reconstruction, achilles, knee, ruptured, recovering, acl, surgeon, meniscus, tendon, injury
blues	reggae, bluegrass, fillmore, rhythm, rockers, ellington, grooves, techno, dnb, hob hawks, leafs, sabres, bruins, tahs, fulham, yotes, rovers, gunners, maroons
origin	ethnicity, ancestry, identity, significance, mythology, identification, protagonists, lineage, lore, retelling nrl, afl, maroons, qld, footy, ashes, wallabies, a-league, premiership, roosters
wing	right-wing, far-left, faction, left-wing, zionist, reactionary, globalist, conservative, extremist, liberal flank, footed, fullback, retake, netting, seat, midfield, fullbacks, guard, mozzarella
weigh	meddle, defer, invest, bathe, reassure, implicated, experts, ponder, expel, summarize weigh-in, weigh-ins, ins, sparring, pre-fight, ufc, bellator, strikeforce, spar, ufcfightpass
worlds	universes, history's, colliding, realms, planets, universe, eras, modes, franchises, environments europeans, olympics, worldcup, commonwealths, wc, commonwealth, championships, european, cwg, paralympics
sessions	comey, rosenstein, recusal, mcgahn, mccabe, recused, recuse, mueller, doj, dhs practices, sess, circuits, drills, weights, interval, camps, trainings, training, workout
track	rmx, compilation, reloaded, hexagon, soundcloud, ep, dnb, bandsintown, tunes, rework racetrack, racin, sx, field, slicks, velodrome, circuit, mtb, race, racing
presents	luts, voyager, housecall, ottaviani, uploaded, balearic, inharmony, derringer, machel, schulz pressies, pressie, advent, decorating, cupcakes, toys, x-mas, sweets, certificates, handmade
AlignCos Top-10	lawrence, marc, morris, op, diamond, carter, dash, cont, bee, norman

Table 7: Top-10 detected words from our method (NN) vs. AlignCos method (last row), for corpus split according to the occupation of the tweet-author. Each word from our method is accompanied by its top-10 neighbors in each of the two corpora (creator vs. sports).



Occupation (creator vs. performer)	
NN	neighbors in each corpus
echo	distortion, echoing, google's, lcd, ibooks, vibe, voice, songbook, audience, roku griffith, park, regents, acjokes, crest, roxy, paramount, trippers, folly, petco
inc	kopel's, acquires, takeover, selects, async, -short, sony, invests, blaqqstarr, tata aimless, caa, phonte, psi, edu, morillo, fuentes, omega, intl, int'l
cont	rec, thru, mang, recs, mi, ul, sr, bsm, ing, tm thku, 😊rt, oth, btw-, muah, 0)777, vry, twd, 🤔rt, wnt
presents	luts, voyager, housecall, ottaviani, uploaded, balearic, inharmony, derringer, machel, schulz morillo, erick, bash, pressies, whalum, pressie, winans, pawty, productions, torry
rebel	kurdish, libyan, jihadist, factions, sunni, jihadi, militant, hamas, daesh, isis ruler, rocker, geek, muse, whore, nerd, madonna's, daydream, gangster, hippie
buck	manziel, clayton, jerry, wiley, cowboys, romo, ambrose, flacco, kidd, mavs bucky, cocker, paperboy, rickie, hefner, mcdowell, roddy, cy, farmer, leadoff
thee	salute, paraphrase, bishop, esv, browning, faulkner, lia, medina, kaysha, atwood shalt, thyself, merciful, ephesians, hahahahahahah, thine, philippians, yesssssss, throne, humbly
chapter	prologue, prc, outlining, novella, pages, scene, heartstopper, cebu, tome, outline bl, tblst, sdmf, doom, grimnest, warhammer, quilt, draculas, dario, crusade
dash	jnr, peppermint, flashes, wop, keef, cappuccino, scotty, hummus, lily, disco skeetv, skee, radio, hbr, snip, twirl, 📺, blip, iheart, krispy
op	hoare, pathologist, shauna, baden-clay, nedrow, barrister, arguedas, protestor, bourque, arias urgentdogsofmiami's, doreenvirtue's, dermatologist, examination, surgeon, physio, intv, ons, nasal, doctor
AlignCos Top-10	vince, todd, dana, watson, norman, marc, jerry, rs, mitch, brooks

Table 8: Top-10 detected words from our method (NN) vs. AlignCos method (last row), for corpus split according to the occupation of the tweet-author. Each word from our method is accompanied by its top-10 neighbors in each of the two corpora (creator vs. performer).

Time of week (weekday vs. weekend)	
NN	neighbors in each corpus
trick	sudoku, sneaky, summat, moonwalk, frighten, rubik's, clicker, smthng, stunt, foam treaters, treater, tricker, trick-or-treating, trick-or-treat, treats, or, neices, trick-or-treaters, kids
cc	citibank, debit, wachovia, credit, barter, visa, waived, payment, pkg, expedia snyder, ecu, rivera, mvc, yankees, clinches, natl, lin, ul, rk
ups	dhl, upping, situps, gowalla, shipment, shipments, fy, lunges, webos, sit-ups budgets, tractor, full-time, dri, radioshack, quik, distribution, fro, cheeseburgers, soulja
recall	recalling, maclaren, stork, cribs, strollers, defective, pedals, tundra, toyota, manufacturer fancy, specify, attribute, recommend, resist, adjust, vary, fwiw, grieve, refrain
rush	stampede, queues, detour, stretch, layover, standstill, congestion, levin, oncoming, braving refusal, jerry, pass, cbs, sellout, sideline, dover, interference, onside, tuscaloosa
bond	etfs, bernanke, insurer, sentencing, trustee, r.i., deficits, rba, hig, funds labor, humphrey, clarke, srk, titanic, fireman, colonel, fx, barney, jessie
pitch	bullpen, clinch, utley, win-win, lidge, interviewed, series, signage, stun, teleconference midfield, half-time, werth, tsn, offside, scoreless, roughing, punts, goal, rockies
lloyd	marv, asher, peter, andre, payton, phillip, bennett, o'connor, neal, wright llyod, jeward, mcelderry, lloyd's, ollie, stace, danyl's, jedwards, afro, olly's
zone	faction, wasteland, emp, vibin, i.e, l.a., constraints, realms, xtreme, jammin endzone, redzone, fumbled, fumbles, interceptions, touchdown, interference, bounds, interception, romo
ref	salary, overturn, statewide, applicants, amendments, position, ordinance, commissioning, nsw, anc offside, capello, burley, mangini, play-off, officiating, roughing, rooney, interference, fumbled
AlignCos Top-10	maine, evan, griffin, terry, sp, aaron, ken, harris, todd, li

Table 9: Top-10 detected words from our method (NN) vs. AlignCos method (last row), for corpus split according to the time of week of the tweet. Each word from our method is accompanied by its top-10 neighbors in each of the two corpora (weekday vs. weekend).

French (2014 vs. 2018)

NN	neighbors in each corpus
malcom	charmed, futurama, desperate, housewives, housewives, simpson, hunter, ferb, smallville, scott dembele, coutinho, mariano, paulinho, rafinha, diakhab, dembelé, dembelé, dembouz, rakitic
rn	en, eb, zn, en., enn, bored, bloquee, same, omfgg, stm fn, rn., dlf, fn., lfi, fhaine, lr, ex-fn, lrem, pcf
boe	bne, bnne, bonne, binne, bonnne, boonne, bone, bnn, bonnee, booonne peiffer, fourcade, svendsen, makarainen, schempp, desthieux, guigonnat, kuzmina, dahlmeier, tarjei
mina	kenza, ibtisssem, bety, ghada, lina, laith, bzf, liya, ana, salom yerry, yerri, paulinho, gomes, mina., alcacer, rakitic, rafinha, dembele, coutinho
smet	smette, smette, tmet, met, spouse, senjaille, stappe, smettent, sdonne, samuse hallyday, laeticia, laura, læticia, vartan, halliday, hallyday., johnny, boudou, laetitia
lr	bdx, dk, poitiers, bx, rouen, caen, amiens, malo, perpi, aix lr, lrem, dlf, lfi, fn, ump, républicains, udi, vb, rn
divergent	tmr, tfios, thg, catching, hunger, mockingjay, fsog, insurgent, allegiant, tobias divergent., diverge, divergentes, diffèrent, convergent, diverger, diamétralement, concordent, opposées., divergences
ia	ya, y', yaura, quya, yavai, yaver, yora, yavait, yia, jconai artificielle, intelligenceartificielle, i.a, ia., intelligence, iot, i.a., artificielle., chatbots, automatisisation
jdr	jdr, hablais, duale, pfpfpfpfp, eso, igual, nadie, déjame, pensar, pelis jdr., warhammer, shadowrun, roleplay, pathfinder, shmup, fangame, dungeon, rp, webcomic
cs	csst, ceest, enpls, wch, tst, cetei, wcch, c, ctei, cetai csgo, rl, pubg, fornite, fortnite, battlerite, faceit, ow, cod, dota
AlignCos Top-10	-l, malcom, maximilien, dna, lr, mina, boe, dias, sierra, giuseppe

Table 10: Top-10 detected words from our method (NN) vs. AlignCos method (last row), for corpus split according to the year of the text. Each word from our method is accompanied by its top-10 neighbors in each of the two corpora (2014 vs. 2018).

Hebrew (2014 vs. 2018)

NN	neighbors in each corpus
באלוהים	באלוהיםםם, נשבעתתתת, באלוהיםםם, אמאלה, באלוהיםם, יאו, יואו, נשבעתתת, נישבעת, נשבעת דקלומי, בביאת, בישוע, באלוהים', באמונות, לאמונה, במשיח, באלוקים, האמונה, אמונה
פיו	הבקייע, בונקר, ראיוס, לזהבי, שלושער, חאמס, ראדי, זהבי, שזהבי, וזהבי, הלוע, מצפוננו, הג'ורה, שפיו, וליבו, מנבל, פיה, הפה, פיהם, פיך,
זהה	יואו, עשיתי, עושהההה, אמלההה, זה, זהההההה, זהההההה, זההההה, זההההה, זהההה, מקבילה, לכמרים, היפותטי, מנוגדת, הדומה, מקביל, לשלך, תואמת, ולבכירים, לציות,
קים	לירז, נוי, שונטל, אוריין, ניקול, לורן, תרין, קטי, הנטל, לין, טראמפ-קים, מקוריאה, קים-טראמפ, בפיוניאנג, הפסגה, לבנגרונד, גונג, בסינגפור, ג'ונג-און, ז'ונג,
בנינו	אחרתתת, החוםםם, ניגמר, יפריד, בסדררר, החוםםם, נגמררר, וויתרת, החוםםם, חיילינו, ילדנו, הרסנו, חיילנו, הלוחמים, ילדינו, ובנינו, בננו, בניכם, לבינם,
בוהן	בהיסטוריה, בכימיה, מתכונת, מבחון, ומבחן, בספרות, באזרחות, בפזיקה, בתנך, בביווגיה, מנתח, תבחן, לבחינת, יבחן, בוחנת, אבחן, בחינת, מציג, מאבחן, שבוחן,
דברי	סבבי, אע, ימסור, ותבוא, תרגיע, טובוש, בנאי, תתקשרי, תדברי, תקשרי, מזמורי, דבריה, שדברי, בדברי, דבריך, ההבל, דבריי, הבלע, ודברי, דיברי,
מאי	לין, לירז, זרבי, בתאל, תרין, רלי, שירז, אוריאן, קורל, שיראל, החודש, מרס, ינואר, בחודש, נובמבר, דצמבר, אוגוסט, אוקטובר, יוני, פברואר,
היתר	לטענת, תמיכת, הפרדת, גיור, האיטור, הפערים, הקרע, רבנות, חילוקי, מקומיות, הוצ, האישורים, המסביב, פולני-יהודי, השוני, היתרים, סיבתי, וההבדל, ההיתר, היתרי,
כול	חידוד, "י", אהובך, ארוסך, חייד, שז, כולכך, ותצאי, אהז, בשאר, ונגמרו, חתיחת, המחורבנים, כשבכל, והטרוריסטים, ואחר, כ-ל, ולכול, כלללל,
AlignCos Top-10	קים, הארי, קרי, מיקי, היומי, אופיר, מעריב, אופק, פרי, באלוהים,

Figure 4: Top-10 detected words from our method (NN) vs. AlignCos method (last row), for corpus split according to the year of the text. Each word from our method is accompanied by its top-10 neighbors in each of the two corpora (2014 vs. 2018).