Short-term meaning shift: an exploratory distributional analysis

Anonymous NAACL submission

Abstract

We investigate diachronic meaning shift that takes place in short periods of time (short-term meaning shift) and in an online community of speakers. We create a small dataset and use it to assess the performance of a standard model for meaning shift detection on short-term meaning shift, and find that this phenomenon poses specific difficulties for models based on the Distributional Hypothesis.

1 Introduction

Semantic change has received increasing attention in empirical Computational Linguistics / NLP in the last few years (**?**). Almost all studies so far have focused on meaning shift in long periods of time –decades to centuries. However, the genesis of meaning shift and the mechanisms that produce it operate at much shorter time spans, ranging from the online agreement on words meaning in dyadic interactions (**?**) to the rapid spread of new meanings in relatively small communities of people (**??**). This paper is, to the best of our knowledge, the first exploration of the latter phenomenon in empirical CL/NLP, that we call *short-term meaning shift*.

We take what seems to us a natural first step, namely, to analyze the behavior of a standard distributional model of semantic change when applied to short-term shift. This kind of model is based on the hypothesis that a change in context of use mirrors a change in meaning. Our results show that a distributional successfully detects most short-term meaning shifts, but that it overgeneralizes, since some contextual changes do not correspond to a meaning shift. We also show that this is a difficulty caused by the nature of short-term meaning shift, and propose to use contextual variability as a means to remedy it.

G: to do: We also should mention that we compare what we find to long-term shift. M: I am not sure I understand what comparison you are referring to

Short-term shift is usually hard to observe in standard language, such as the language of books or news, which has been the focus of long-term studies (**??**), since it takes a long time for a new meaning to be widely accepted in the standard language. We therefore focus on the language produced in an online community of speakers, in which the adoption of new meanings happens at a much faster pace (**??**).

Unlike studies of long-term meaning shift, we cannot rely on material previously gathered by linguists or lexicographers for analysis or evaluation. For this reason, we create a small dataset to enable empirical analysis of short-term shift analysis and to allow comparison in future studies.[1]

2 Related Work

Several methods have been proposed to investigate long-term meaning shift: common to all of them is the computation of time-related distributional representations for words in the vocabulary, and the sequential comparison of such representations in order to detect a drop in self-similarity, usually interpreted as a shift in meaning.

Among the most widely used techniques are Latent Semantic Analysis (**??**), Topic Modeling (**??**), and simple co-occurence matrices of target words and context terms (**??**). More recently, researchers have used word embeddings computed using the skip-gram model by **?**. Since embeddings computed in different semantic spaces are not directly comparable, time related representation are usually made comparable either by aligning different semantic spaces (**???**) or by initializing the em-

[1]Data and code will be made available upon publication.

sample	time bin	million tokens
Reddit_{13}	2013	∼900
LiverpoolFC_{13}	2011–13	8.5
LiverpoolFC_{17}	2017	11.9

Table 1: Time bin and size of the datasets.

beddings at t+1 using those computed at t (**????**). We adopt the latter methodology.

G: to do: Remove redundancy with the intro. I put this in the intro cause I think we need to make it more prominent M: I commented out the last paragraph, and slightly modified the one you added at the end of the intro

3 Experimental Setup

Data. We exploit user-generated language from an online forum of football fans, namely, the r/LiverpoolFC subreddit, one of the many communities hosted by the Reddit platform.[2]M: I would remove this link, it is useless We focus on a short period of eight years, between 2011 and 2017. In order to enable a clearer observation of short-term meaning shift, we define two non-consecutive time bins: the first one (t_1) contains data from 2011–2013 and the second one (t_2) from 2017.[3] We also use a large sample of community-independent language for the initialization of the word vectors, namely, a random crawl from Reddit in 2013.[4] Table 1 shows the size of each sample.

Model. In the method proposed by **?**, word embeddings for the first time bin t_1 are initialized randomly; then, given a sequence of time-related samples, embeddings for t_i are initialized using the embeddings of t_{i-1} and further updated. If at t_i the word is used in the same contexts as in t_{i-1}, its embedding will only be marginally updated, whereas a major change in the context of use will lead to a stronger update of the embedding. The model makes embeddings across time bins directly comparable.

We implement the following steps:[5] First, we create word embeddings with the large sample Reddit_{13}, to obtain meaning representations that are community-independent.[6] M: Useless footnote, we say about the random initialization before We then use these embeddings to initialize those in LiverpoolFC_{13}, update the vectors on this sample, and thus obtain embeddings for time t_1. This step adapts the general embeddings from Reddit_{13} to the LiverpoolFC community. Finally, we initialize the word embeddings for LiverpoolFC_{17} with those of t_1, train on this sample and get embeddings for t_2.

The vocabulary is defined as the intersection of the vocabularies of the three samples (Reddit_{13}, LiverpoolFC_{13}, LiverpoolFC_{17}), and includes 157k words. For Reddit_{13}, we include only words that occur at least 20 times in the sample, so as to ensure meaningful representations for each word, while for the other two samples we do not use any frequency threshold: Since the embeddings used for the initialization of LiverpoolFC_{13} encode community-independent meanings, if a word doesn't occur in LiverpoolFC_{13} its representation will simply be as in Reddit_{13}, which reflects the idea that if a word is not used in a community, then its meaning is not altered within that community. We train with standard skip-gram parameters (**?**): window 5, learning rate 0.01, embedding dimension 200, hierarchical softmax.

Evaluation dataset. The dataset includes approximately 97 words. They were annotated by 26 members of the r/LiverpoolFC subreddit, that is, community members without linguistic background, because domain knowledge is needed for this task. For each word, subjects read contexts of use from the two time bins and, for simplicity, were asked to make a binary decision about whether there was a change in meaning. However, semantic shift is arguably a graded notion, so we aggregate the annotations into a graded *semantic shift index*, ranging from 0 (no shift) to 1 (shift) depending on how many subjects spotted semantic change. This index is shown in the y-axis of Figure 1. Further methodological details regarding data collection are described in Appendix (n).

4 Sources of Meaning Shift

We identify two main sources of shift in our data. G: To do: which are common to long-term / known, which are new / particular to short-term?

[2] `https://www.reddit.com`.

[3] These choices ensure that the samples in these two time bins are approximately of the same size – see Table 1. The r/LiverpoolFC subreddit exists since 2009, but very little content was produced in 2009–2010.

[4] We used the Python package Praw for downloading the data, `https://pypi.python.org/pypi/praw`.

[5] We implement the model using the `gensim` library.

[6] Reddit_{13} embeddings are initialized randomly.

M: I am not able to say right now. I guess fig lang related ones are in common, while memes ones are not. However, I have not enough knowledge of the literature to back it up.The first are analogical processes related to figurative language, in particular metaphor and metonymy. G: to do: break down following into metaphor - explanation - example(s), then metonymy - explanation - example(s) M: I see your point, but I am bit in trouble when I need to assign a label: for example, is 'highlighter' an example of metaphor or metonymy? Maybe we could just talk about fig lang in particular metaphoric usage of words - which leads to broadening or loosening of the original meaning of the word, and metonymy - when a highly salient characteristic of an entity is used to refer to it as a whole. Among these cases are, for example, 'highlighter', which occurs in sentences like *'we are playing with the **highlighter** today'* or *'what's up with the hate for this kit? This is great, ten times better than the **highlighter**'*, used to talk about a kit in a colour similar to that of a highlighter; or 'lean', in *'I hope a **lean** comes soon!'* and *'Somebody with speed...make a signing... Cuz I need a **lean**'*, due to players typically leaning on a Liverpool symbol when posing for a photo right after signing for the club. Particularly explanatory is the 'F5' example shown (in chronological order) in Table 2. While 'F5' is initially used with its common usage of shortcut for refreshing a page (line 1), it then starts to denote the act of refreshing in order to get the latest news about the possible transfer of a new player to LiverpoolFC (2). This use catches on and many redditors use it to express their tension while waiting for good news (3-5),[7] though not all members are aware of the new meaning of the word (6). After the player finally signed for the team, someone leaves the 'F5 squad' (7), and after a while, another member recalls the period in which the word was used (8).

The second main source are memes. In this case, fans start to use a word to make jokes or as a motto, and the new usage quickly spreads within the community. It is the case of 'monitoring': after Liverpool monitored some promising players for some months without in the end signing any of them, fans started to use the word in fixed sentences (*'**Monitoring** intensifies!'*) which are perceived as sarcastic by the community, or to make jokes like in *'I've switched my focus to **monitoring** Jessica Alba'* G: To do: choose clearer example (I don't know who J. Alba is...) M: mmm, I think the example is ok, she's super famous. The meme became so pervasive that some users expressed their disappointment, like in *'I'm getting tired of the "monitoring" jokes'*.

[7]Here the semantic change is accompanied by a change in the part of speech, and 'F5' becomes a denominal verb.

1.	after losing the F5 key on my keyboard...	18 Jun
2.	F5 tapping is so intense now. I want him	28 Jun
3.	Don't think about it too much, man. Just F5	1 Jul
4.	just woke up and thought it was f5 time	3 Jul
5.	this was a happy f5	13 Jul
6.	what is an F5?	13 Jul
7.	I'm leaving the f5 squad for now	5 Aug
8.	I made this during the f5 madness in the sub	6 Sept

Table 2: Examples of use of 'F5'. On the third column is the date of the post: the short time span from the first to the last example gives the idea of how quick is the meaning shift process considered in this work.

5 Results and Analysis

Our initial hypothesis, common to previous work, is that meaning shift is mirrored by a change in context of usage, which should be captured by an increased in cosine distance between the time-related vector representations of a word, in a graded fashion (more semantic change, more contextual change). The results of our experiment support this hypothesis: We find a positive correlation between cosine distance and semantic shift in our dataset (Pearson's r= 0.49, $p < 0.001$). This result indicates that the model is indeed able to capture the majority of the cases of semantic shift in our data, as it can be observed in Figure 1, in which the examples presented in Section 4 are highlighted.

Although the general tendency is in line with our expectations, we also find systematic deviations. First, *false positives*, that is, words that do not undergo semantic shift despite showing relatively large differences in context between t_1 and t_2 (red ellipsis in Figure 1; shift index=0, cosine distance>0.25). Manual inspection reveals that most of these "errors" are due to a referential effect: words are used almost exclusively to refer to a specific person or event, and so the context of use is narrowed down with respect to t_1. For instance, 'stubborn' is almost always used to talk about a coach who was not there in 2013 but only in 2017; 'entourage', for the entourage of one of the stars of the team; 'independence' for the political events of

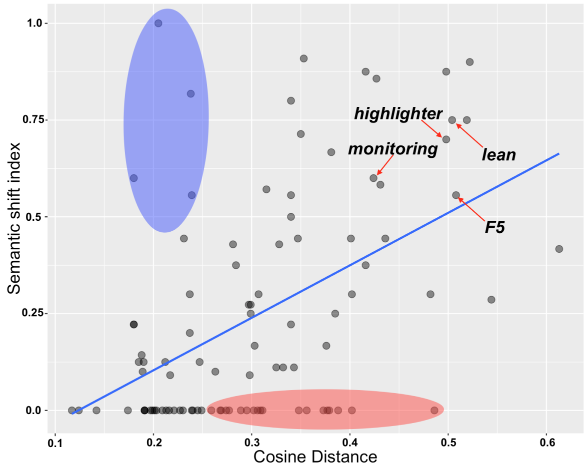

Figure 1: Semantic shift index vs. cosine distance for all words in the evaluation dataset (Pearson's r = 0.49, $p < 0.001$). Red ellipsis indicates false positives, blue ellipsis false negatives.

Catalonia (Spain). In all these cases, the meaning of the word stays the same, despite the change in context. In line with the Distributional Hypothesis, the model spots the change, but it is not sensitive to its nature. Long-term shift is not as sensitive to changes of referential nature like these because embeddings are aggregated over a larger and more varied number of occurrences.

A smaller, but consistent group is that of *false negatives*, words that undergo semantic shift but are not captured by the model (blue ellipsis; shift index>0.5, cosine distance<0.25). G: To do: move this to the ling analysis section, and here only link? M: don't know, I see it more here maybe. Raq? These are cases of *extended* metaphor (**?**), that is, cases in which the metaphor is developed throughout the whole text produced by an author. Also in this case, the model "is right", in the sense that indeed the local context of the target words does not change in t_2. For instance, 'pharaoh' is the nickname of an Egyptian player who joined Liverpool in 2017 and is used in sentences like *'approved by our new **Pharaoh** Tutankhamun'*, *'our dear Egyptian **Pharaoh**, let's hope he becomes a God'*, and so on. Similarly, 'shovel', occurs in sentences like *'welcome aboard, here is your **shovel**'*, *'you boys know how to **shovel** coal'*: the team is seen as a train that is running through the season, and every supporter is asked to give its contribution, depicted as the act of shoving coal into the train boiler. Despite the metaphoric usage, the local context of these words is similar to the literal one, and so the model does not spot the meaning shift. We expect this to happen in long-term shift models, too, but we are not aware of results confirming this.

6 Contextual variability.

From the analysis it emerges that the main issue for distributional models when dealing with short-term shift is contextual change due to referential aspects (false positives). We expect that in referential cases the context of use will be *narrower* than for words with actual semantic shift, because they are specific to one person or event. Hence, using a measure of *contextual variability* should help spot false positives. Here we test this hypothesis. We define contextual variability as follows: for a target word, we create a vector for each of its contexts in t_2 by averaging the embeddings of the words occurring in it, and define variability as the average pairwise cosine distance between context vectors.[8] We then test whether contextual variability has explanatory power over cosine distance by fitting a linear regression model with these two variables as predictors and semantic shift index as dependent variable.[9] The results indicate that these two aspects are indeed complementary (contextual variability: β= 0.47, $p < 0.001$, cosine distance: β= 0.40, $p < 0.001$, adjusted R^2=0.44). While both shift words and referential cases change context of use in t_2, context variability captures the fact that only in referential cases words occur in a restricted set of contexts. The scatterplot 2 shows this effect visually. This result can inform future models of short-term meaning shift. In future work, we plan to investigate more in depth the interplay between variability, cosine and semantic shift, both in short- and long-term meaning shift.M: I'd move this to Conclusions

7 Conclusion

The goal of this preliminary study was to bring to the attention of the NLP community short-term meaning shift, an under-researched problem in the field. We hope that it will spark further research into a phenomenon which, besides being of theoretical interest, has potential practical implications for NLP downstream tasks concerned with user-generated text, as modeling how words meaning

[8] We consider the context as the five words occurring on the left and on the right of the target word.

[9] Contextual variability and cosine distance are not correlated in our data (Pearson's r= 0.18,$p > 0.05$).

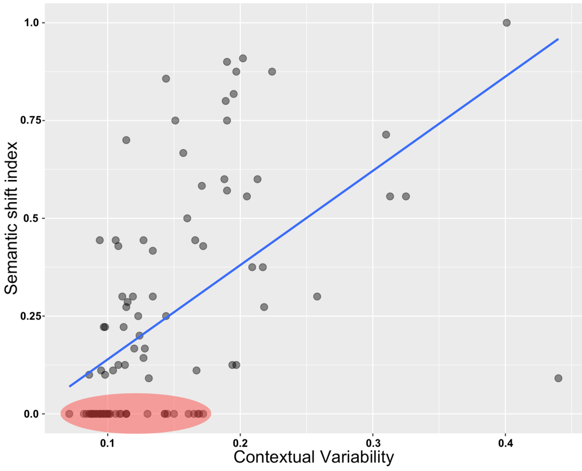

Figure 2: Semantic shift index vs. context variability for all words in the evaluation dataset (Pearson's $r = 0.55$, $p < 0.001$). Red ellipses indicates the referential cases which are incorrectly assigned high cosine distance values (false positives in the paper).

rapidly change in communities would allow to better understand what their members say.