DialectGram: Detecting Dialectal Variation at Multiple Geographic Resolutions

Hang Jiang*
Symbolic Systems
hjian42@stanford.edu

Haoshen Hong*
Computer Science
haoshen@stanford.edu

Yuxing Chen*
Symbolic Systems
yxchen28@stanford.edu

Vivek Kulkarni
Computer Science
viveksck@stanford.edu

Abstract

Several computational models have been developed to detect and analyze dialect variation in recent years. Most of these models assume a predefined set of geographical regions over which they detect and analyze dialectal variation. However, dialect variation occurs at multiple levels of geographic resolution ranging from cities within a state, states within a country, and between countries across continents. In this work, we propose a model that enables detection of dialectal variation at multiple levels of geographic resolution obviating the need for a-priori definition of the resolution level. Our method DIALECTGRAM, learns dialect-sensitive word embeddings while being agnostic of the geographic resolution. Specifically it only requires one-time training and enables analysis of dialectal variation at a chosen resolution post-hoc – a significant departure from prior models which need to be retrained whenever the pre-defined set of regions changes. Furthermore, DIALECTGRAM explicitly models senses thus enabling one to estimate the proportion of each sense usage in any given region. Finally, we quantitatively evaluate our model against other baselines on a new evaluation dataset `DialectSim` (in English) and show that DIALECTGRAM can effectively model linguistic variation.

1 Introduction

Studying regional variation of language is central to the field of sociolinguistics. Traditional approaches (Labov, 1980; Milroy, 1992; Tagliamonte, 2006; Wolfram and Schilling, 2015) focus on rigorous manual analysis of linguistic data collected through time-consuming and expensive surveys and questionnaires. The evolution of the Internet and social media now enables studying linguistic variation at a scale thus overcoming some of the scalability challenges faced by survey based methods. Consequently, computational methods to detect and analyze geographic variation in language have been proposed (Eisenstein et al., 2010, 2011, 2014; Bamman et al., 2014; Kulkarni et al., 2015b)

However, most prior work suffers from three limitations: First, previous models (Kulkarni et al., 2015b) such as Frequency Model, Syntactic Model, and `GEODIST` all rely on pre-defined regional classes to model linguistic changes (an exception is (Eisenstein et al., 2010) which focuses on lexical variation). The use of pre-defined regional classes limits the flexibility of these baseline models because dialect changes can be observed at various geographic resolutions. Second, previous models do not explicitly model the sense distribution of each word. In this work, we address these limitations by proposing a model DIALECTGRAM that enables analysis at multiple geographic resolutions while explicitly modeling word senses (see Figures 1 - 4). Given a corpus which can be associated with geographical regions, DialectGram first induces the number of senses for each word using a non-parametric Bayesian model (Bartunov et al., 2016). This step requires no apriori knowledge of the geographic resolution[1]. Having inferred the senses of each word, we show how to detect and analyze dialectal variation at any chosen geographic resolution by clustering usages in any given region based on their sense usage.

To summarize, our contributions are:

- **Multi-resolution Model:** We introduce DIALECTGRAM, a method to study the geographic variation in language across multiple

*Equal contribution.

[1]The only requirement is that the corpus be geo-tagged so that analysis can be conducted post-hoc at any desired resolution.

levels of resolution without assuming knowledge of the geographical resolution apriori.

- **Explicit Sense modeling:** DIALECTGRAM predicts how likely each sense of a word is used in a context thus enabling a more precise modeling of linguistic change.
- **Corpus and Validation Set:** We build a new English Twitter corpus `Geo-Tweets2019` for training dialect-sensitive word embeddings. Furthermore, we construct a new validation set `DialectSim` for evaluating the quality of English region-specific word embeddings between UK and USA.

2 Related Work

Linguistic variation. In the past, sociologists and linguists have been studying linguistic change by designing experiments to manually collect data (Labov, 1980; Milroy, 1992) and conducting variation analysis (Tagliamonte, 2006). Several works (Eisenstein et al., 2010; Gulordava and Baroni, 2011; Kim et al., 2014; Jatowt and Duh, 2014; Kulkarni et al., 2015a,b; Kenter et al., 2015; Gonçalves and Sánchez, 2016; Donoso and Sanchez, 2017; Lucy and Mendelsohn, 2018; Shoemark et al., 2019) have used different computational models to study dialect variations with respective to geography, gender, and time.

Eisenstein et al. (2010) is one of the first to tackle the linguistic variation problem with computational models. They design a multi-level generative model that uses latent topic and geographic variables to analyze lexical variation in English. This latent variable model is able to generate an author's geographic location based on the author's text. To quantitatively evaluate the models, they compute the physical distance between the prediction and the true location. Similarly, Gonçalves and Sánchez (2016) apply K-means method to cluster the geographic lexical superdialects assuming a list of pre-defined set of words that are known to demonstrate lexical variation. This was followed by Gonçalves and Sánchez (2016) who propose two metrics to calculate the linguistic distance between geographic regions. That is, instead of using the physical distance between the predicted and the true location, they compute cosine similarities or Jensen-Shannon Divergence (JSD) to evaluate the model quantitatively.

Recently, Kulkarni et al. (2015b) building on the work of (Bamman et al., 2014) propose a word embeddings based model `GEODIST` model for robustly modeling dialectal variation and focuses on capturing semantic changes between dialects. Nevertheless, a pre-defined set of regions is required for the model to update region-specific embeddings. For instance, Kulkarni et al. (2015b) assume that English exhibits dialectal variation between the US and UK, and train the network to learn two sets of word embeddings for the two regions. However, a model trained using this data cannot be used to analyze dialectal variation across states or any other level of resolution without a re-training from scratch. To learn how English changes within each state, Kulkarni et al. (2015b) would need to tag each US tweet with a state name and train the model again. Moreover, the model does not explicitly capture senses of a word but only learns region specific embeddings.

Word Sense Disambiguation. The problem of detecting dialectal variants of a word can be viewed broadly in terms of word sense induction where the different word senses can roughly correspond to usages in different regions. For instance, the word *pants* usually refer to *underwear* in the US versus *trousers* in the UK, suggesting two senses for *pants*. Consequently, we discuss the most relevant work on word sense induction as well. Reisinger and Mooney (2010) is the first paper that modifies the single *prototype* vector space model to obtain multi-sense word embeddings with average cluster vectors as prototypes. Many works (Huang et al., 2012; Neelakantan et al., 2014; Tian et al., 2014; Chen et al., 2014) are later dedicated to combine Skip-gram, clustering algorithm, and linguistic knowledge to learn word senses and embeddings jointly. Bartunov et al. (2016) adopt a non-parametric Bayesian approach and propose the Adaptive Skip-gram (AdaGram) model, which is able to induce word senses without assuming any fixed number of prototypes. As we will see in the following sections, we build on precisely this approach to model regional variation.

3 Data

3.1 Geo-Tweets2019 Corpus

We create a new corpus, `Geo-Tweets2019`, which consists of English tweets[2] during April and May in 2019 from the United States and the United Kingdom. Each tweet includes the user ID, the

[2]We use the Tweepy toolkit.

Word	US Meaning	UK Meaning
flat	smooth and even; without marked lumps or indentations	apartment
flyover	flypast, ceremonial aircraft flight	elevated road section
pants	trousers	underwear
lift	elevator	raise
football	soccer	American football

Table 1: Examples of words that have different meanings in American and British English

published time, the geographic location, and tweet text. We have around 2M tweets from the US and 1M from the UK. We preprocessed the tweets with the tweet tokenizer from Eisenstein et al., 2010 and regular expressions. Finally, we filtered out URL's, emojis, and other irregular uses of English to shrink the size of vocabulary and to facilitate the training of word vectors. Statistics can be seen in Table 2.

Number	US	UK	Total
tweet	2,075,394	1,088,232	3,163,626
token	41,637,107	22,012,953	63,650,060
term	865,784	469,570	1,167,790

Table 2: Statistics of Geo-Tweets2019

3.2 DialectSim Validation Set

To evaluate the models, we construct a new validation set `DialectSim`, which comprises of words with same or shifted meanings in the US and the UK. To build this validation set, we first crawled a list of words that show different meanings from the Wikipedia page[3] and pick 341 words that appear more than 20 times in our corpus in the UK and the US. Table 1 presents three examples in the dataset. In order to generate balanced positive and negative samples, we sample another 341 negative examples randomly from our `Geo-Tweets2019` dataset. A minimum frequency of 20 is also used for negative sampling. These negative cases were manually verified by each of the three authors independently. Finally, we split the dataset into training set with 511 samples (75%) and testing set with 171 samples (25%).

4 Models

4.1 Baseline Models

Frequency Model. One baseline method to detect whether there are significant changes between usage in two regions is to count the occurrence of a word in the US and the UK tweets. We have implemented this Frequency Model as described in Kulkarni et al. (2015b).

Syntactic Model. A more nuances approach compared to the frequency based approach is to detect change in syntactical roles across regions. The Syntactic Model (Kulkarni et al., 2015b) takes Part-of-Speech (POS) tag into consideration as well. More specifically, if a word is used equally frequently in both countries, but the their POS usages are different, then we consider the meaning of two words as different between two countries. We use the CMU ARK Twitter Part-of-Speech Tagger[4] for POS tagging.

GEODIST (Skip-gram) Model. The main idea of `GEODIST` model (which can detect semantic changes) (Kulkarni et al., 2015b) is to learn region-specific word embeddings and use boot-strapping to estimate confidence scores on detected changes. Instead of learning a single vector to represent a word, this model aims to jointly learn a global embedding $\delta_{\text{MAIN}}(w)$ as well as (multiple) differential embeddings $\delta_{r_i}(w)$ for each word w in the vocabulary with $R = (r_1, r_2, \dots)$ geographical regions exactly as described in (Bamman et al., 2014). In particular, the region-specific embedding is defined as the sum of the global embedding and the differential embedding for that region: $\phi_{r_i}(w) = \delta_{\text{MAIN}(w)} + \delta_{r_i}(w)$. The objective function is to minimize the negative log-likelihood of the context word given the center word conditioned on the region. We use stochastic gradient descent method (Bottou, 1991) to update the model parameters. We implement our own `GEODIST` model in PyTorch.

4.2 DialectGram Model

We construct a new model for detecting dialectal changes which we called DIALECTGRAM (Dialectal Adaptive Skip-gram). The model first learns multi-sense word embeddings using Ada-

[3] https://en.wikipedia.org/wiki/Lists_of_words_having_different_meanings_in_American_and_British_English

[4] http://www.cs.cmu.edu/~ark/TweetNLP/

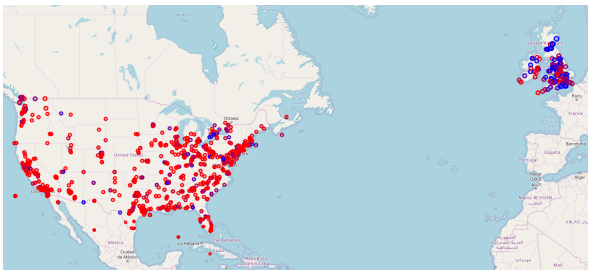

Figure 1: Dialectal variation of *gas* across countries. Tweets that contain *gas* with predicted sense "gaseous substance" are illustrated as blue circles; tweets that contain *gas* with predicted sense "gasoline" are plotted as red circles.

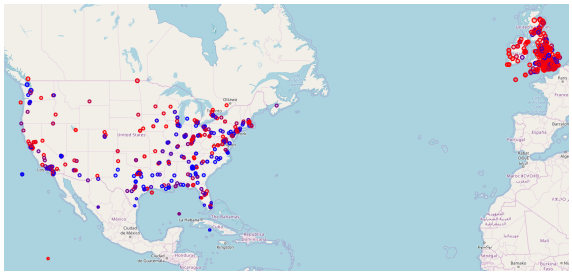

Figure 2: Dialectal variation of *flat* across countries. Tweets that contain *flat* with predicted sense "apartment" are illustrated as red circles; tweets that contain *flat* with predicted sense "smooth and even" are plotted as blue circles.

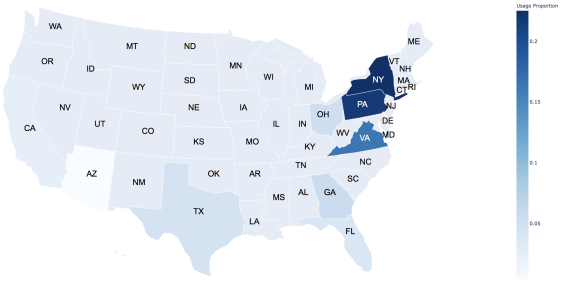

Figure 3: Dialectal variation of *buffalo* across US states. Here we show for each state, the proportion of sense 1 usage (*Buffalo city*) in blue. Grey indicates that the state contains no tweet using the word *buffalo* in our corpus.

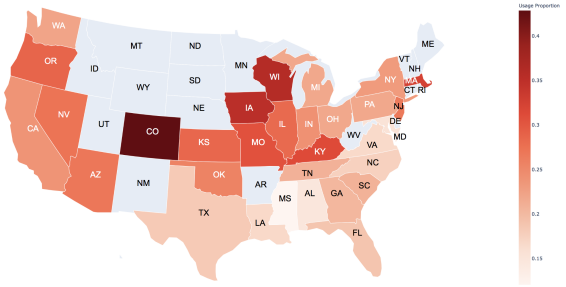

Figure 4: Dialectal variation of *pop* across US states. Here we show for each state, the proportion of sense 2 usage (*soft drink, soda*) in red. Grey indicates that the state contains no tweet using the word *pop* in our corpus.

gram (Bartunov et al., 2016) through training on the region-agnostic corpus. Once sense specific embeddings are obtained, based on the chosen resolution the model composes region-specific word embeddings by taking a weighted average of sense embeddings. At last, the model calculates the distance between region-specific word embeddings of the same word to determine whether a significant change exists. Our method is described succinctly in Algorithm 1.

Compared to the GEODIST model which needs predefined geographic label to update the region-specific embeddings, DIALECTGRAM learns multi-sense word embeddings on our dataset without any knowledge of the underlying regions.For instance, DialectGram automatically induces and learns the two senses of the word *flat* which could mean an *apartment* or *level land* corresponding to usages in the UK and US respectively.

Implementation details We train our model on our Geo-Tweets2019 corpus to learn word sense embeddings using the Julia implementation of AdaGram[5] and then implement the inference algorithm in Python. To obtain a word's region-specific embedding in a place, we first use DIALECTGRAM to predict the dominant sense for the word in each tweet from a region and use weighted average of the sense embeddings as the region-specific word embedding e_r. We use the fol-

Algorithm 1 Use DIALECTGRAM to Compose Region-specific Embeddings

Input: w word
Output: e_r weighted region embedding for w

```
Load the trained DialectGram model
Build Index_r on Corpus from region r
for s, p ∈ GetSensePriors(w) do
    S_c[s] ← 0, S_p[s] ← p ▷ Note: S_c is sense counts, S_p is sense priors
end for
for all c ∈ GetContexts(w) do
    s ← Disambiguate(w, c)
    S_c[s] ← S_c[s] + 1
end for
e_r ← GetWeightedVector(S_c, S_p)
```

[5] https://github.com/sbos/AdaGram.jl

lowing hyper-parameter settings: min_freq = 20, window_size = 10, dimension = 100, maximum_prototype = 30, $\alpha = 0.1$, epoch = 1, sense_threshold = $1e-17$. It is worth noting that a large α (the underlying Dirichlet process) may lead to too many senses for some words and a small α, on the contrary, results in too few senses.

To measure the significance of the dialectal change, Kulkarni et al. (2015b) propose an unsupervised method to detect words with statistically significant meaning changes. However, given that we have access to the humanly curated DialectSim dataset, we evaluate the models on the list of annotated words using a simple thresh-holding model (where the thresh-hold parameter is learned from training data). Specifically, We evaluate both Skip-gram models (i.e. GEODIST and DIALECTGRAM) by calculating the Manhattan distance[6] between a word's region-specific embeddings[7].

5 Results

5.1 Qualitative Analysis

We investigate the words that GEODIST model predicts to have a significant dialectal change between the two regions. For example, the word *mate* is one of the top 20 words in our vocabulary if we sort the vocabulary by the Manhattan distance between the US and the UK embeddings from high to low. However, words like *draft* are predicted to have different regional meanings but not labelled as "significant" in DialectSim. We further discuss this issue in section 5.2.3.

We select some words with significantly different meanings between the UK and the US. In our DIALECTGRAM model, we select the most frequent 2 senses, which usually account for more than 99% usage variation of a word, and plot a heat map on world map.

The word maps in Figure [1, 2] suggest that the usage of *gas* and *flat* are different in the UK and in the US. *Gas* is used commonly as petrol and related to gas station in the US, but in the UK, *gas* usually refers to air and natural gas. *Flat* could refer to *apartment* but in the US this meaning is not as common as in the UK. The same model can also be used at a different resolution level (across US states). For example, given the word *buffalo*, we show the most dominant senses where *Buffalo City* (in blue) and the *buffalo sauce* sense (in white). Similarly for the word *pop*, we observe that the Midwest area and the Pacific Northwest are more reddish, indicating people are more likely to use the word for *soft drink, soda*, while people in other areas like to use it to describe a certain type of music – *pop music* [8].

5.2 Quantitative Results

Our training corpus Geo-Tweets2019 has over three million tweets from US and UK. However, we still observed that micro-level analyses at a resolution lower than the state level required more data samples. Therefore, we only present the country-level and state-level analysis here (note that we do not need to train the model to learn embeddings again when we change resolutions for our analyses).

For each model, we defined a score function that takes in one word and return a real number denoting its difference in meanings between the UK and the US. We fit a simple threshold model that maximizes the accuracy on training set. Then we test the model performance on testing set. The results are shown in Table 4.

5.2.1 Frequency Model

We observed that Frequency Model is more sensitive to word difference between two countries: *football* in the UK is same as *soccer* in the US, causing an imbalanced frequency of term *football* between both countries. However, it can not detect some semantic changes of words if the semantic change preserves frequency for both countries: *flat* has similar frequency in both countries, despite the fact that *flat* could mean *apartment* in the UK, whereas this usage is uncommon in the US. This model does not suffer from an over-fitting problem, because the model is fairly simple and the parameter space is quite small. However the Frequency model is susceptible to a high false positive rate.

5.2.2 Syntactic Model

Syntactic Model performs the worst among all the models. It still gets slightly higher precision than

[6] We tried euclidean and cosine distance as well, but use Manhattan distance since it yielded the best results out of the three metrics.

[7] Our models, validation set and code are available at: https://github.com/yuxingch/DialectGram.

[8] We normalized the data points by filtering out states where the number of tweets is less than 15 since a small number of data points can suffer from high variance.

word	sense 1 neighbors	sense 2 neighbors
gas	industrial, masks, electric	car, station, bus
flat	kitchen, shower, window	shoes, problems, temperatures
buffalo	syracuse, hutchinson	chicken, fries, seafood
subway	starbucks, restaurant, mcdonalds	1mph, commercial, 5kmh

Table 3: Neighbors of sense embeddings for selected words. This shows DIALECTGRAM is able to learn semantic variations of words.

Model	Acc	Prec	Recall	F1
Frequency	0.5600	0.5600	0.5887	0.5568
Syntactic	0.5263	0.5714	0.4828	0.5233
GEODIST	0.6432	**0.7424**	0.5810	0.6518
DIALECTGRAM	**0.6667**	0.6837	**0.6438**	**0.6632**

Table 4: Test performance. Acc, Prec means accuracy and precision. DIALECTGRAM has better accuracy, recall, and F1 score than GEODIST.

the Frequency Model on test set because it gets some dialectal syntactic changes correct. There are two reasons for its bad performance. First, it is limited by the performance of POS Tagger. Second many word sense changes do not alter POS tags. For example, *pants* refers to *underwear* in the UK while it refers to *jeans* in the US, and both of them are nouns.

5.2.3 GEODIST Model

As mentioned in Section 5.1, GEODIST model is able to detect dialect changes. The accuracy on the test set beats the previous two baseline models (0.6432 versus 0.5600 and 0.5263), as shown in Table 4. It also outperforms the baseline models in terms of precision and F1 score. In fact, GEODIST model has the highest precision among all models, including the DIALECTGRAM model that will be discussed in the next section. We also notice that the recall on the test set is the lowest. The high precision with low recall indicates that for those changes that GEODIST model is very conservative and misses some words that actually have significant dialectal changes. For example, the difference between the two region-specific embeddings of the word *pants* is predicted to be not significant, while *pants* does have different meanings in the UK and the US (Table 1).

5.2.4 DialectGram Model

DialectGram outperforms the GEODIST model in accuracy, recall, and F1 score. However, its precision is lower than that of the GEODIST and Frequency Model. However, this is already impressive given the fact that DialectGram does not require pre-determined geographic labels and enables analysis at different geographic resolutions post-hoc (after the model is trained). One reason for DIALECTGRAM's lower performance in precision compared to GEODIST model is that it overestimates the number of senses (learning senses that overlap). For example the word *gas* in Table 3, we sometimes have an additional sense characterized by words such as *air, house, pipe*. This sense seems to be a mix of sense 1, gaseous substance, and sense 2, gasoline. The average number of senses is controlled by α which we pick based on the model's performance on the training set, but we acknowledge that smarter search strategies for α could be employed.

6 Conclusion

In this work, we proposed a novel method to detect linguistic variations on multiple resolution levels. In our new approach, we use DIALECTGRAM to train multiple sense embeddings on region-agnostic data, compose region-specific word embeddings, and determines whether there is a significant dialectal variation across regions for a word. In contrast to baseline models, DIALECTGRAM does not rely on the region-labels for training multi-sense word embeddings. The use of region-agnostic data allows DIALECTGRAM to conduct multi-resolution analysis with one-time training. We also construct Geo-Tweets2019, a new corpus from online Twitter users in the UK and US for training word embeddings. To validate our work, we also contribute a new validation set DialectSim for explicitly measuring the performance of our models in detecting the linguistic variations between the US and the UK. This validation set allows for more precise comparison between our method (DIALECTGRAM) and previous methods including Frequency Model, Syntactic Model, and GEODIST model. On DialectSim, our method achieves better per-

formance than the previous models in accuracy, recall, and F1 score. Through linguistic analysis, we also found that DIALECTGRAM model learns rich linguistic changes between British and American English. Finally, we conclude by noting the method can be easily extended to temporal or analysis of language at multi-resolution levels.

Acknowledgments

We would like to thank Cindy Wang, Christopher Potts, and anonymous reviewers, who gave precious advice and comments to our paper. We would also like to thank Symbolic Systems Program at Stanford University for funding our research through Grants for Education And Research (GEAR).

References

David Bamman, Chris Dyer, and Noah A Smith. 2014. Distributed representations of geographically situated language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834.

Sergey Bartunov, Dmitry Kondrashkin, Anton Osokin, and Dmitry Vetrov. 2016. Breaking sticks and ambiguities with adaptive skip-gram. In *Artificial Intelligence and Statistics*, pages 130–138.

Léon Bottou. 1991. Stochastic gradient learning in neural networks. *Proceedings of Neuro-Nımes*, 91(8):12.

Xinxiong Chen, Zhiyuan Liu, and Maosong Sun. 2014. A unified model for word sense representation and disambiguation. In *EMNLP*.

Gonzalo Donoso and David Sanchez. 2017. Dialectometric analysis of language variation in twitter. *arXiv preprint arXiv:1702.06777*.

Jacob Eisenstein, Brendan O'Connor, Noah A Smith, and Eric P Xing. 2010. A latent variable model for geographic lexical variation. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, pages 1277–1287. Association for Computational Linguistics.

Jacob Eisenstein, Brendan O'Connor, Noah A Smith, and Eric P Xing. 2014. Diffusion of lexical change in social media. *PloS one*, 9(11):e113114.

Jacob Eisenstein, Noah A Smith, and Eric P Xing. 2011. Discovering sociolinguistic associations with structured sparsity. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1365–1374. Association for Computational Linguistics.

Bruno Gonçalves and David Sánchez. 2016. Learning about spanish dialects through twitter. *Revista Internacional de Lingüística Iberoamericana*, pages 65–75.

Kristina Gulordava and Marco Baroni. 2011. A distributional similarity approach to the detection of semantic change in the google books ngram corpus. In *Proceedings of the GEMS 2011 workshop on geometrical models of natural language semantics*, pages 67–71.

Eric H. Huang, Richard Socher, Christopher D. Manning, and Andrew Y. Ng. 2012. Improving word representations via global context and multiple word prototypes. In *ACL*.

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*, pages 229–238. IEEE Press.

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*, 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. *arXiv preprint arXiv:1405.3515*.

Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015a. 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.

Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2015b. Freshman or fresher? quantifying the geographic variation of internet language. *arXiv preprint arXiv:1510.06786*.

William Labov. 1980. *Locating language in time and space*. Academic Press New York.

Li Lucy and Julia Mendelsohn. 2018. Using sentiment induction to understand variation in gendered online communities. *arXiv preprint arXiv:1811.07061*.

James Milroy. 1992. *Linguistic variation and change: On the historical sociolinguistics of English*. B. Blackwell.

Arvind Neelakantan, Jeevan Shankar, Alexandre Passos, and Andrew McCallum. 2014. Efficient non-parametric estimation of multiple embeddings per word in vector space. In *EMNLP*.

Joseph Reisinger and Raymond J. Mooney. 2010. Multi-prototype vector-space models of word meaning. In *HLT-NAACL*.

Philippa Shoemark, Farhana Ferdousi Liza, Dong Nguyen, Scott A Hale, and Barbara McGillivray. 2019. Room to glo: A systematic comparison of semantic change detection approaches with word embeddings.

Sali A Tagliamonte. 2006. *Analysing sociolinguistic variation*. Cambridge University Press.

Fei Tian, Hanjun Dai, Jiang Bian, Bin Gao, Rui Zhang, Enhong Chen, and Tie-Yan Liu. 2014. A probabilistic model for learning multi-prototype word embeddings. In *COLING*.

Walt Wolfram and Natalie Schilling. 2015. *American English: dialects and variation*, volume 25. John Wiley & Sons.