HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

"This is lit, fam":

Diachronic word embeddings and classifying semantic change

University of Helsinki

Faculty of Arts

Department of Languages

English philology

Undergraduate Thesis

May 2020

Teemu Pöyhönen

014696400

Supervisor: Joseph Flanagan

Contents

1 Introduction

Investigating semantic change, its regularity and causes, have been key processes of examination in linguistic research. Language adapts to the needs of its users, and studies generally suggest that changes in word meaning occur systematically. Hence, linguists have been able to both categorize incidents and describe the change as a process. Moreover, it is important to maintain the research and to analyze the manner in which semantic change occurs, as this leads us to understand its developments and the needs of users of any language.

Most of the semantic change research in the field of diachronic linguistics has focused on changes developing over centuries or multiple decades. A well-known example of this type of semantic shift is the change in the word form “gay” – though in the modern day it is synonymous with “homosexual,” before the 19th century it used to contain the senses “merry” or “joyful.” Moreover, Bloomfield (1933) has introduced categories for semantic change, such as *narrowing/broadening* and *degeneration/elevation*, which have been in wide use. In this way, with enough data, it is relatively accessible for linguists to research this type of gradual change and attempt to categorize these changes.

Inspecting semantic change in our contemporary language use, however, has been challenging, as the attempt to detect linguistic patterns occurring in a short time frame may yield questionable results. With this, word embeddings have emerged as a promising methodology for distributional semantics through the increasing intersection of humanities and digital resources. This quantitative methodology has prompted computational methods to be employable by linguists, consequently increasing knowledge about the quantitative aspect of language change. With word embeddings, using collocational patterns to analyze the similarity between word forms may be suitable for quantifying semantic relationships in corpora. Specifically, this methodology may be applicable to study short term language change in our contemporary language.

The aim of this paper is to examine whether word embeddings can be used effectively to provide evidence for semantic shifts, and specifically, how we may operationalize traditional categories of semantic change in terms of word embeddings. This study begins by using a sampled corpus to train word embeddings. It will then analyze the words *swipe, lit, fam, sick* and *toxic* in terms of whether any semantic change can be detected within them. These changes will then be categorized based on Bloomfield’s categories of semantic change. For this, in section two, I will give a background to semantic change and introduce some of the types of semantic change proposed by Bloomfield: *narrowing/broadening, elevation/degeneration, substitiution,* and *metaphor*. Then, I will discuss the background to distributional semantics, and the linguistic theory which word embeddings are based upon. In section three, I will present the data and methods developed for the word embedding model in this paper, and then attempt to analyze the data. Finally, I will discuss the results which seem to suggest that word embeddings seem to be an effective tool to detect semantic change, yet some of the categories of semantic change are problematic when employed in a context of short periods.

2 Background Literature

2.1 Semantic Change

Historical linguistics examines changes in the meaning of a word form. In this way, either the change in the quantity of senses a word form contains or the changes in the connotation of a word is viewed as semantic change. Traditionally, considerable amount of work gone into the research of semantic change consists of categorizing and documenting various types of change (Kutuzov et al, 2018). For this, Bloomfield (1933) defines semantic shifts as "innovations which change the lexical meaning rather than the grammatical function of a form." (p. 425) In this way, semantic change differs from other kinds of language change in that inherent features of word forms are not under inspection, and it is moreover the fundamental lexicon and the relationships between lexemes that are examined for change.

A predominant way of both discussing and researching semantic change has been to compare cognates from different languages and periods. In this instance, Bloomfield discusses how the "comparison of related languages shows different meanings in forms that we feel justified in viewing as cognate." For example, the Old English word *mete* 'food' translates more recently to the English word *meat*. Additionally, Bloomfield gives a comparison of how the form for *chin* agrees in meaning with German *kinn*, Dutch *kin,* but Gothic *kinnus* and Old Norse *kinn* represent the meaning *cheek.* (p. 425) In this way, observing how cognates differ and develop in meaning within language families can be concluded as semantic change.

These kinds of changes that have been argued by Bloomfield are nevertheless changes which are linguistic in nature. In contrast, Hamilton, et al. (2016) argue for a distinction between the semantic change occurring due to linguistic drifts and cultural factors. It is essential to define "whether changes are more cultural or linguistic in nature, a distinction that is essential for work in the digital humanities and historical linguists" (p. 1). Changes occurring due to cultural factors, such as new technologies, cause semantic change—"such as the change in the meaning of cell ("prison cell", "cell phone")" As one examines semantic change, the distinction emerges as we move away from studying cognates and language families, and the occurring semantic change yields from a non-linguistic cause. More recently, Gulordava and Baroni (2011), showed "the word sleep acquiring more negative connotations related to sleep disorders, when comparing its 1960s contexts to 1990s contexts."

(Kutuzov et al. 2018, p. 3.) In this way, changes in word connotations occur often in shorter periods and due to cultural shifts. Moreover, these changes occur within a single language and are less associated with cognate developments, and more with the perceptions of speakers or change in the speakers' culture.

2.2 Categories in Semantic Change

Based on Bloomfield's definition of semantic shifts, he proposed and hence after is widely referred to for his nine classes of semantic shifts. Similar mechanisms that represent the opposite cases of each other have been grouped into the same section. It should be noted, that these are not all of the types by Bloomfield, as only the relevant ones have been selected for discussion here.

2.2.1 Narrowing-Broadening

In *narrowing* or *broadening,* the meaning of a word changes in range, so it is in this way appropriate in different amount contexts than before. As an example of *broadening*, *dog* used to mean a specific breed of dogs, yet today it is used to refer to all breeds of dogs. The aforementioned example of *mete* 'food' into *meat* is then an example of *narrowing*.

2.2.2 Elevation-Degeneration

A classic example of this type of change is the *elevation* of the Old English word *cniht* 'boy, servant' into *knight*. In this way, the meaning is elevated into a more positive one in the mind of a speaker, or the referrent is valued more by a speaker. In a more modern example, we have seen *queer* change its meaning from a slur into a term used by people in the LGBT community to identify themselves.

2.2.3 Metaphor

Metaphor concerns change in which the meaning of a word is extended. In this way, the word form may after the change contain more senses since a new sense is added to the word form. An important factor for the senses is that the new sense must in some way be related to the original sense of the word. More specifically, the connection is often metaphorical in nature. For example, the word *grasp* used to only mean holding or gripping physically, yet today the word is also used in more abstract contexts, where the word is a close synonym to *understand,* e.g. "to grasp a concept."

2.2.4 Substitution

While *substitution* is not part of Bloomfield's nine types of semantic change, it has been introduced by Stern (1931) to describe the semantic change which reflects a change in the referent. In this way, "*substitution* describes a change that has a non-linguistic cause, namely that of technological process" (Kutuzov et. al 2018, p. 2). This is directly related to the distinction between a cultural and linguistic change argued by Hamilton, et al. (2016). In this way, *substitution* is a category of cultural type of semantic change. Namely, the word *car* shifted its meaning from non-motorized vehicles after the introduction of the automobile. (Kutuzov, et al. 2018 p. 3). In this sense, *substitution* can differ in terms of categorization, as its cause is non-linguistic, as opposed to the previously mentioned traditional types of semantic change which have a linguistic cause.

2.3 Distributional Semantics

Distributional semantics is based on the Distributional Hypothesis by Harris (1954). This states that "similarity in meaning results in similarity of linguistic distribution." (Boleda, 2020). Firth (1957) also follows this by discussing a distributional semantic approach to words and their collocational study. In this framework, word meaning (and connotation) emerges from the word's collocational patterns which in turn provide us an "arbitrary definition of the word" (p. 26). Boleda gives us the examples of *postdoc* and *student*, where these words would be "used in similar contexts, as in *a poor _, the _ struggled through the deadline* (Boleda & Herbelot 2016, p. 623). Thus, the words *postdoc* and *student* could be argued to have similar meanings.

Due to working within a framework of distributional semantics, we are lead to an assumption that "diachronic changes in collocational patterning should be taken to reflect the semantic change of a construction" (Hilpert, 2008, p. 181). In this way, if a word co-occurs with different words in different periods, it can be assumed that the meaning of the word has changed. This assumes a contextual and usage-based position on semantics, present in this paper. Moreover, this view "aligns well with the assumptions underlying the distributional semantic approach" (Kutuzov 2018, p. 3).

These frameworks are rather appropriate as one employs quantitative research methods. In fact, "semantic shifts are often reflected in large corpora through change in the context of the word which is undergoing shift, as measured by co-occurring words. It is thus natural to try to detect semantic shifts in a data-driven way" (Kutuzov, et al. 2018.) In this way, if we are able to observe a change in a word's collocational patterns, we may assume that the word's usage has been changed as well. In this way, "a change in context of use mirrors a change in meaning, which can be regaded as a special case of the Distributional Hypothesis." (Boleda 2020, p. 217). Moreover, as long as we make sure the corpora are sampled into different time intervals, they lend themselves ideally to this data-driven way of detecting semantic shifts.

The research using quantitive methods within distributional semantics has mostly shown that this framework can model semantic change, but the field has not so far given much attention to furthering our understanding of semantic change (Boleda 2020, p. 218.) Kutuzov et al. (2018) argue a similar lack of detailed analysis, in which for example, one could do "sub-classification of types of semantic shifts (broadening, narrowing, etc.)" (p. 10) This could be attributed to the fact that the emerging field is still relatively new, and it is perhaps still justifying its methods.

3 Data & Methodology

3.1 Raw data

Year	Sentences	Types		Tokens	
		Before Training	After Training	Before Training	After Training
2006	376447	168189	154111	15677214	12356285
2007	2181115	411312	351465	78320310	61715616
2008	6150133	757770	612279	210901880	165941756
2009	16605157	1690683	1235638	582564993	457732308
2010	23914755	2219691	1537855	787189992	618122117
2011	23958826	2363089	1603743	748887302	587966194
2012	23908192	2529138	1699340	732757492	576641116
2013	23900319	2715379	1781634	714132534	563358671
2014	23878087	3005985	1966453	727950133	577508929
2015	23834818	3069442	2003634	718466355	572949466
2016	23816493	8733430	2470213	765591380	610351357
2017	23787385	2997398	1991538	724815948	579327847
2018	25666596	3002826	1977149	737597428	589333269
2019	23640602	2704144	1800232	630855704	503848088

Table 1. The numbers for raw word types, word tokens and sentences. The counts are given before and after training the model, as the minimum count and downsampling parameters affect these counts.

The data used in this paper is a collection of Reddit comments from 2006 to 2019. First, all comments were downloaded from Pushshift, a big-data storage and analytics project. Then, a sampled corpus was created from these comments. This was done by first by taking a random portion of two million comments from each month of all the years. Lastly, all of the comments were combined into text files that contain comments for the entire year. In this way, each year contains approximately 24 million comments, 7 billion tokens and 2 million word types. As an exception, years before 2010 do not contain up to two million comments per month, so the years 2006 to 2009 contain less than 24 million comments.

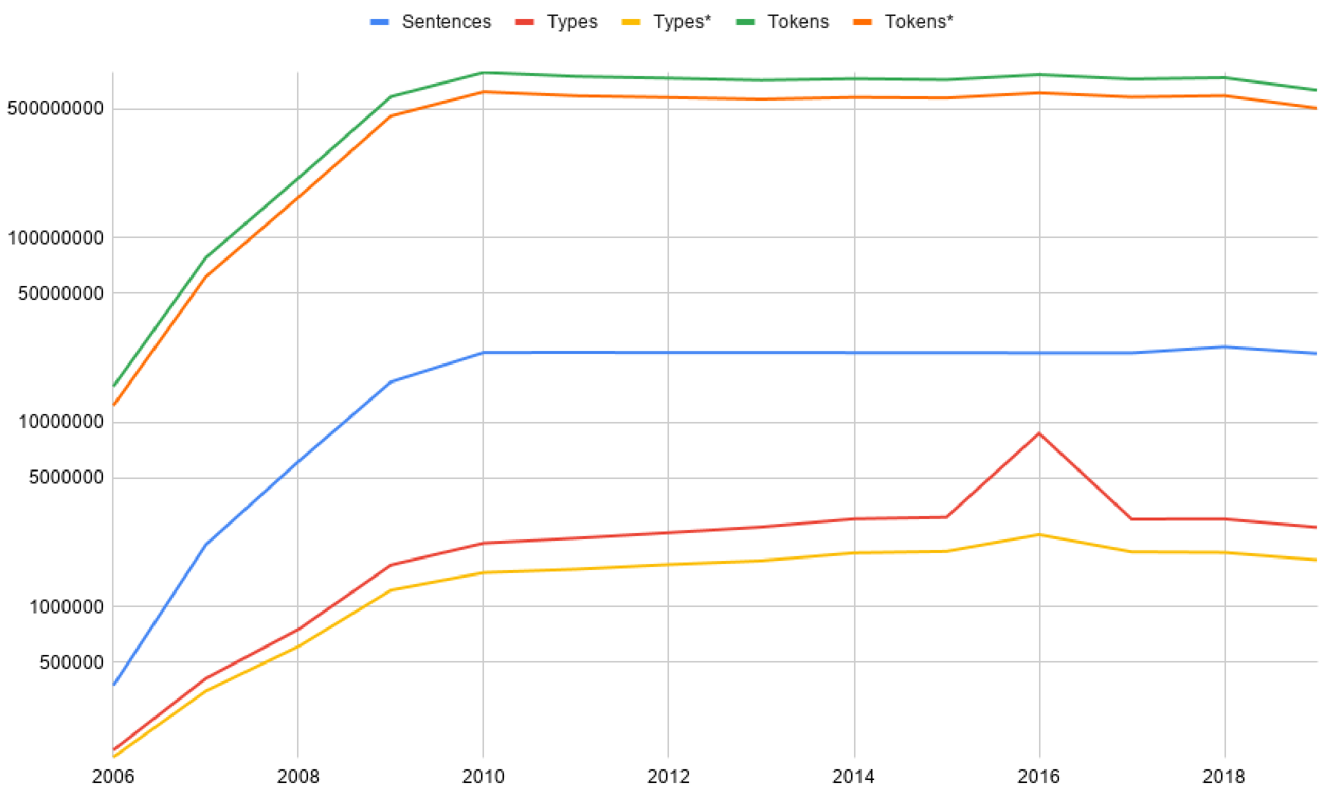

Figure 1. The y-axis is displayed in logarithmic scale to better visualize the change of frequency. The counts marked with an asterisk (*) represent counts after training the model.

3.2 The Reddit demographic

The language data may not be representative of an average language speaker, and we should assume that the model will learn certain types of biases. According to a survey done by Barthlel et al. (2016), a Reddit user is twice as likely to be male than female. Additionally, 64% of the users are aged between 18 and 29. Moreover, 58.4% of the users are located in the United States. In this sense, the average user of Reddit could be described as white, young and male.

While we should be cautious when making generalizations, certain analysis can still be made with the context of a certain demographic in mind. In addition, issues of toxicity in Reddit (and online forums in general) have been raised. These concerns contribute contextual information to our view of how we should interpret the data. Specifically, we have to keep in mind when analyzing the results that the language is not representative of an average speaker, and, more importantly, the issue of toxicity may be relevant when discussing positive or negative connotations of words.

3.3 Word Embeddings

Word embeddings are based on the concept that we take “large amounts of text as input and, through an abstraction mechanism, producing a distributional model, akin to a lexicon, with semantic representations in the form of vectors.” (Boleda 2020, p. 214) Broadly, word embeddings are a statistical method for learning distributional representations of words, or in other words, they are vector representations of particular words. For this, word2vec is a word embedding model which learns embeddings from a corpus and converts the words into mathematical representations in a high dimensional space.

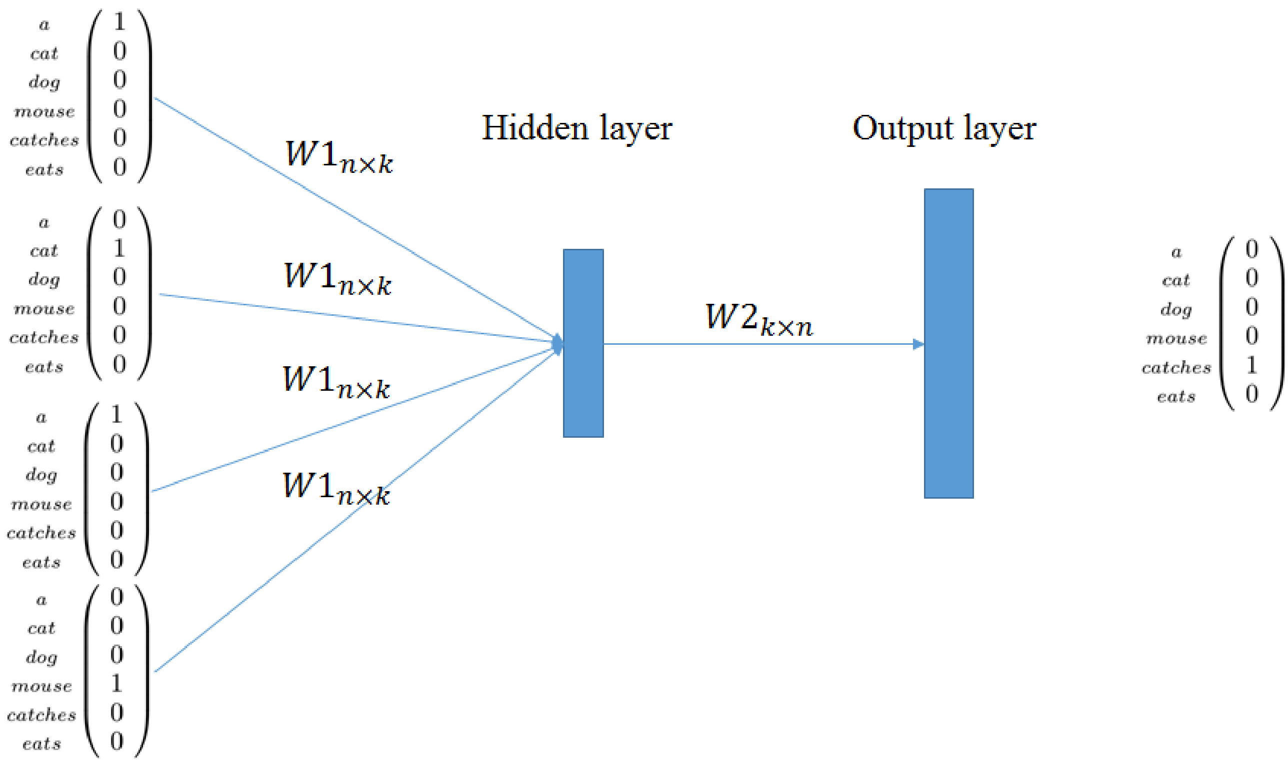

Figure 2. Continuous-bag-of-words (CBOW) model. The neural network uses the context of each word in the corpus to attempt to predict the target word which is *catches* in this case.

This paper employs the continuous bag of words (CBOW) technique, which attempts to predict a word based on the surrounding context. Moreover, CBOW is a neural network that implicitly learns the regularities and distributions, and there is no explicit association feature between pairs of words (Ferrone & Zanzotto, 2017). For example, given the sentence s1 of the corpus [s1: a cat catches a mouse, s2: a dog eats a mouse, s3: a dog catches a cat], the network has to predict *catches* given its context (see Figure 2, Ferrone & Zanzotto, 2017)

Based on the distributional semantics framework, similar words should have similar contexts, hence similar meanings. In this way, “semantic relations can be modeled as geometric relations” (Boleda 2020, p. 215) The distance, or angle, between word vectors produces a cosine similarity score which we may interpret as the similarity between the words. In other words, we may calculate similar words in a word embedding model by comparing their cosine distances within the vector space. By comparing these similarity scores, we may see the degree of which certain senses and contexts appear with the target word.

In reality, these geometrical models are highly dimensional, often several hundred in dimension. For us to be able to visualize these models and vector angles, we project the models into two dimensions. In this way, we may visualize semantic change by observing the vector spaces once they have been projected into two dimensions. Below is an example of the potential of this approach. The target words (*gay, broadcast* and *awful)* and their most similar counterparts reveal changes in meaning when vector spaces from different periods are compared. In this example, we first find all similar words of a target word over the considered time-points. Then, we compute the embeddings of these words on the most recent time-point, as well as the target word’s embeddings for all time-points.

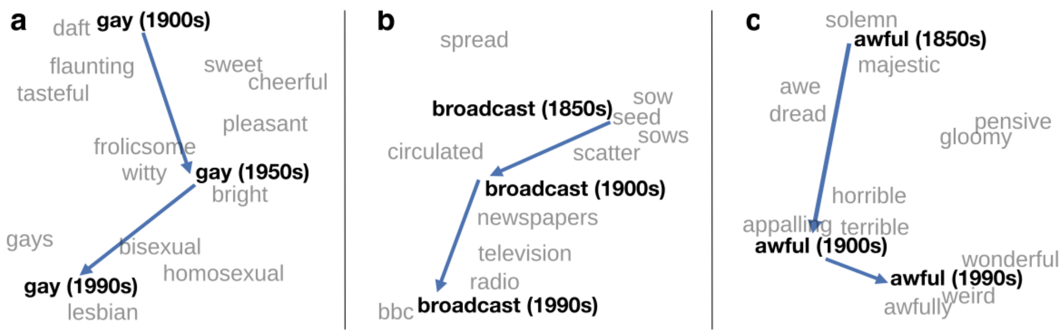

Figure 3. Two-dimensional visualization of semantic change in English using vectors. Figure adapted from Hamilton et al. (2016)

A traditional word embedding method may not consider any time intervals. Since a word embedding model represents a corpus, it represents word collocational patterns of a single time interval. Because of this, we have “To be able to compare embeddings across time, their vector spaces corresponding to different time periods have to be aligned.” (Smith et al. 2017). This arises the alignment problem, as embeddings are always aligned differently, and hence cannot immediately be compared. The solution to this is borrowed from Temporal Word Embeddings with a Compass (TWEC), which is a model created by Di Carlo, et al. (2019). In this model, we use the entire corpus to train an atemporal embedding as a “compass” which is used as a reference for training the other embeddings that are based on sampled corpora of each year. In this way, the “angle” of all these embeddings will be rotated in the same way, and can thus be compared.

3.4 Data Preparation

Each comment is lowercased and tokenized using gensim.utils.simple_preprocess. This process keeps stopwords and most frequent tokens. While most applications discard words such as *the,* the Word2vec algorithm implements a subsampling function, which calculates the probability of keeping the word in a training process. Additionally, each word, that does not meet the minimum count of two occurrences in a corpus, is discarded in the training process.

3.5 Methods

The word embedding model in this paper uses TWEC by Di Carlo, et al. This can be accessed from their GitHub repository. TWEC uses the gensim library, which produces gensim word2vec objects. For training, we set the word2vec objects to have a dimensionality of 300. This was decided due to the relatively large size of the data, as well as due to the intent of capturing both syntactic information as well as some topic modeling information. Each object was trained over ten iterations, with window size and negative sampling being both five.

In this study, we will look at the following words: *swipe, fam, noob, lit, sick* and *toxic.* For each target word, we take the ten most similar words for each year. With these neighboring words, we take their most recent (2019) position on the vector space, as well as the position of the target word for each year. Then, we transform these highly dimensional positions using principal component analysis (PCA), constructing a two-dimensional vector space representation consisting of each of the neighboring words and the target words for each of the years. With this two-dimensional vector space, we may look for evidence whether any semantic change has occurred for the target word by examining how its position changes for each year within the vector space representation.

This paper also employs a statistical method for calculating vector similarity. More specifically, cosine similarity is employed as the technique for comparing word similarity. With this, the most similar words for a target word are used as the motivation for finding the different senses of the target word. Then, three similar neighboring words are selected to represent a specific context a target word is used in. With these words, the similarity scores between the target word and each of the neighboring words are calculated for each year, and the similarity scores of similar contexts are taken to calculate the average similarity score to broadly represent that sense.

4 Results & Analysis

4.1 Swipe

Swipe shows a rapid change. Before 2009, *swipe* is more commonly associated with *card, debit, credit, jab* and *stab*. From 2010 forwards, the word is moving away from these senses,and is moving closer to *rotate* and *tapping*. *Swipe* has seen a change from being predominantly associated with either physically swiping at something (as an attack) or swiping a card (*debit, credit).* With the rise of user interface design, the word has rather quickly gained a sense of often describing a user interaction.

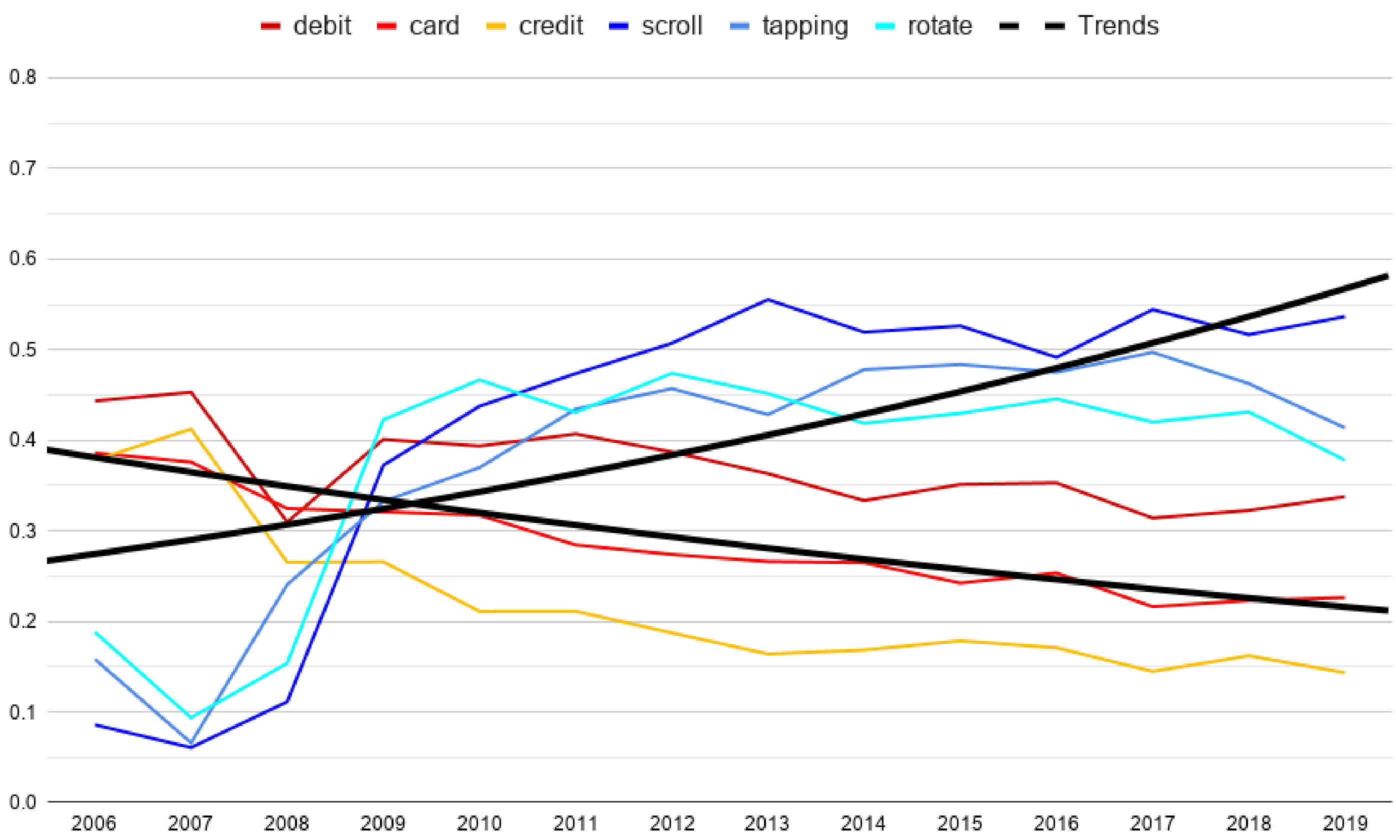

Figure 4. Similarity scores for *swipe* and the exponential trends of the averages of the senses. The increasing trend represents the sense of user interface interaction and the decreasing trend represents the sense of “paying by a card.”

The words *debit, credit* and *card* should be most representative of the word *swipe* in contexts where, for example, a person swipes their card to pay for something. After filtering other forms of *swipe* (such as *swipes, swiping, swiped),* these words were the most similar ones to *swipe* between 2006 and 2010. For example, the average similarity between *swipe* and *debit, credit* and *card* is 0.4 at the beginning, though it decreases to 0.2 by the end of the timeline. Similarly, *tapping, scroll* and *rotate* were selected to represent the context of the word, in which *swipe* is used to mean an interaction in a user interface. The average similarity is quite low in the beginning (less than 0.2 for the first three years), though is steadily above 0.4 from 2012 forwards.

With these results, we can see that there is an emerging sense for the word *swipe*. Specifically, the sense refering to a user interface action does not seem to exist before 2008, as the similarity of the sums of similar contexts is less than 0.2, whereas most recently it is close to 0.6. With this, it would seem that there is a *metaphorical* extension of meaning. In other words, the sense referring to a user interface action, e.g. making a finger gesture on a mobile phone screen, can be seen as a kind of swiping motion and hence the new sense is closely related to the original sense of the word. It remains to be seen whether this semantic change would lead to *narrowing*. So far the change is rather recent, and the other sense of motion is still commonly used.

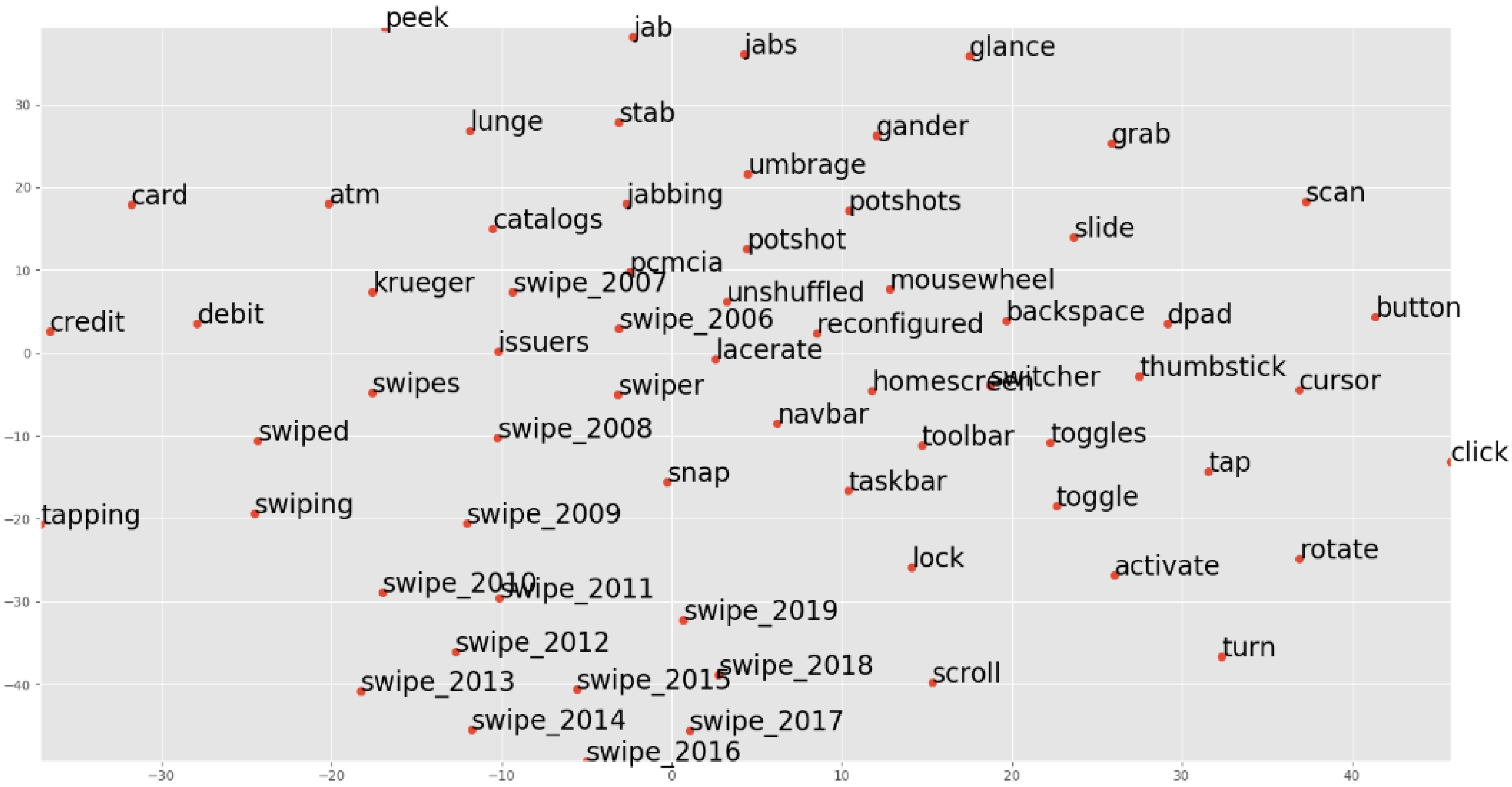

Figure 5. Two-dimensional representation of the vector space for the word *swipe.*

The projected two-dimensional vector space gives us some evidence of the new sense. Since the most similar senses for *swipe* after 2012 refer to the user interaction sense, due to the methodology, most of the words in the vector space are related to that sense, e.g. *backspace, toolbar, taskbar, scroll, lock, activate*. Nevertheless, we see *swipe* move away from the words *card, credit, debit, atm* as well as from such words as *jab, stab, lunge*.

4.2 Fam

The word *fam* has two definitions in many online dictionaries, such as Merriam-Webster: (1) informal abbreviation of *family* and (2) a slang term for a close friend. In the beginning of the timeline, up until 2012, most similar words to *fam* seem to be either informal family member terms such as *sis, hubby. fiancée, mum, stepdad,* or the word *family* (*fam* in this context used as an abbreviation). After this, we can see the more contemporary use of *fam,* the use of referring to close friends and not family members, become the most common context. In this way, *fam* shows consistent and rapid change away from (informal) terms of traditional family members and towards the popular slang it is used today.

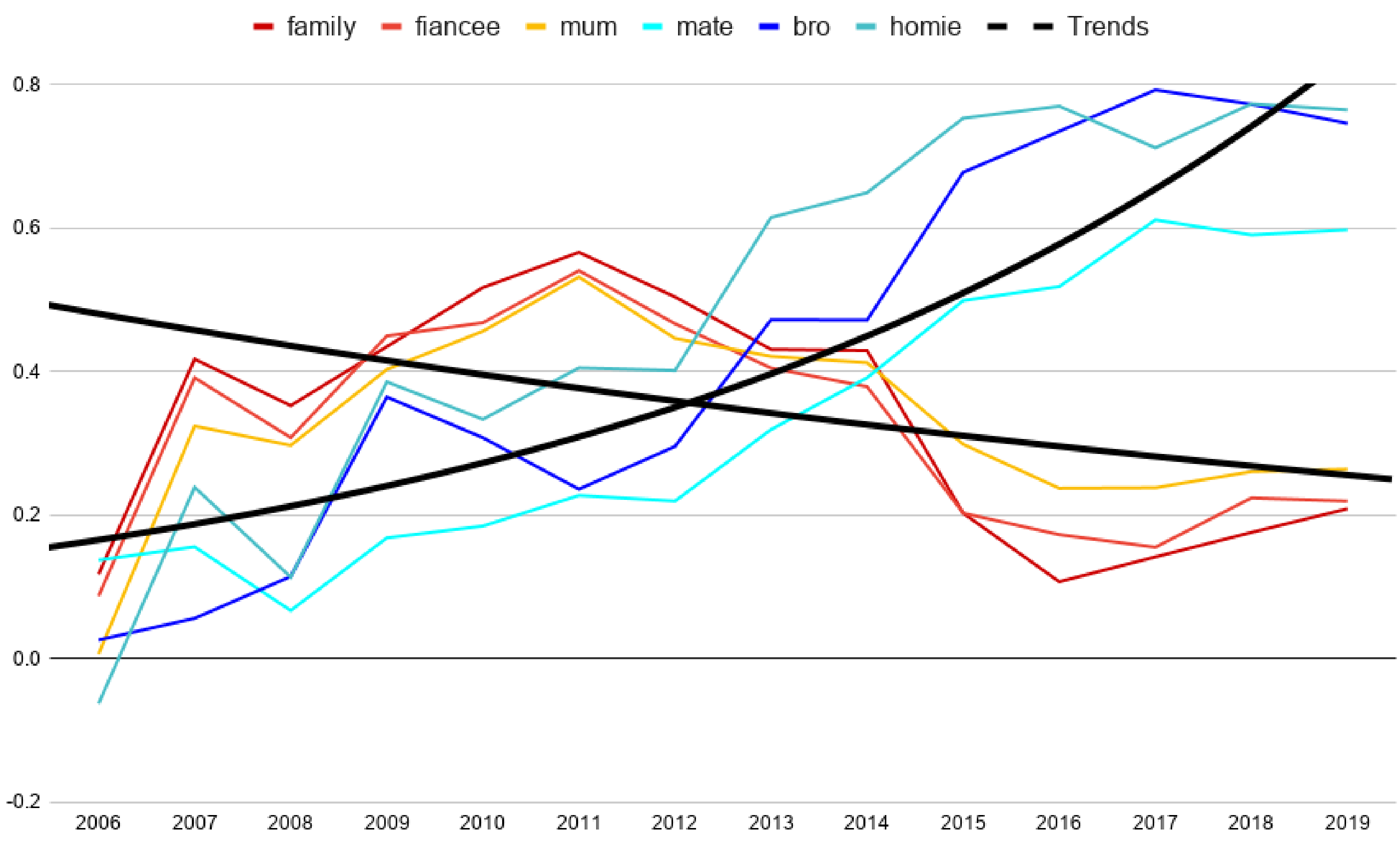

Figure 6. Similarity scores for *fam* and the exponential trends of the averages of the senses. The increasing trend represents the sense of “close friend” and the decreasing trend represents the sense of “family.”

Throughout the years 2008 and 2012, the most similar words to *fam* are words such as *hubby, family, sis, gf, stepmom, stepdad* and *fiancee.* From 2013 forwards, there is a quick change into the most similar words being e.g. *homies, bruh, brutha, brotha, homeboy* and *dawg*. As an exception, the first two years produce results that are seemingly random abbreviations. This is most likely due to *fam* being extremely infrequent in these models. With this, the average similarity of the first group of words starts rather high at 0.5, though declines to 0.2 by the end of the timeline. On the other hand, the second average score for the sense of *homie* increases from 0.1 to 0.7 by the end of the timeline.

It could be argued that *fam* has undergone *broadening*. Since the original sense referred to one's own family, and the new sense to one's close friends, the word now refers to people more generally: whether they are actual family or close friends. Nevertheless, it remains to be seen whether the original meaning will persist, as the similarity scores for the original sense are decreasing quite rapidly after 2014. Moreover, there the new sense for *fam* is appropriated from African American Vernacular English (AAVE) (Merriam-Webster, 2017), which the current paper will discuss in the discussion. Also, it should be noted that the similarity scores for the context of a family are rather low in the beginning of the timeline as well.

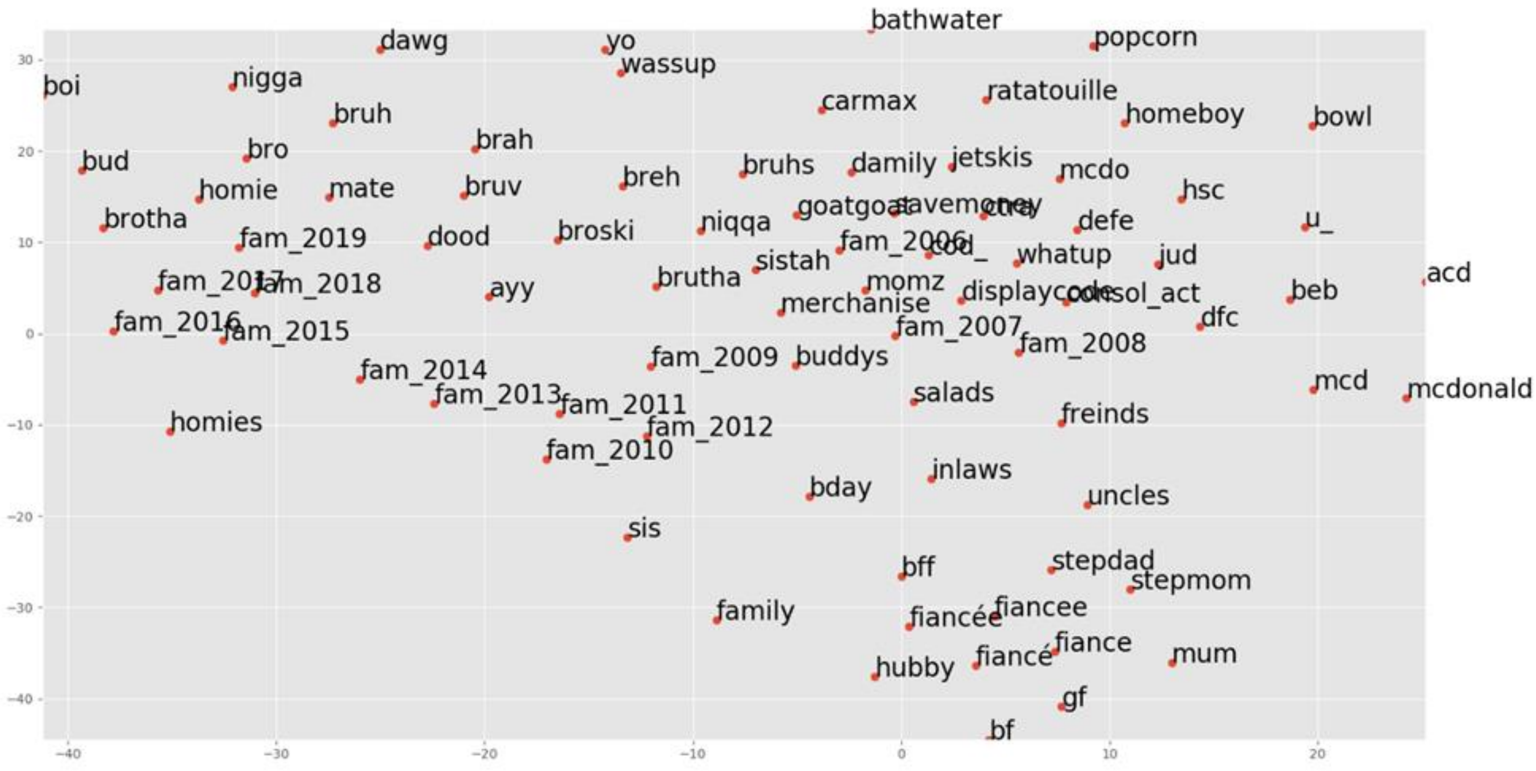

Figure 7. Two-dimensional representation of the vector space of the word *fam.*

The projected vector space for *fam* and its neighboring words seem to be rather representative of the analysis so far. There is a rather clear path away from *family, stepdad, fiancee, stepmom, stepdad* and towards *homie, bro, mate, bruh, bud*. As with some of the seemingly random words, such as *jetskis, defe, jud, hsc* and so on, they are the most similar words to *fam* in the beginning of the timeline. This is most likely due to *fam* being a rather infrequent token throughout 2006 to 2007, and the results are most likely a consequence of this.

4.3 Noob

The word *noob* seems to show change, in which the first four to five years show a considerable shift away from more negative senses, such as *newfag, dumbass, fucktard, moron* and *tard*. The nine most recent years (2010 to 2019) show a consistent shift towards more neutral senses, such as *novice* and *beginner*. Most recently, the most similar words to noob are other variants of the word, such as *newb, newb* and *newbie.* Hence the most similar words to *noob* that are slurs and negative senses were chosen *(fucktard, moron, stupid*) to represent the negative context, and similarly, the most similar words to noob that are relatively neutral were chosen to represent the neutral context.

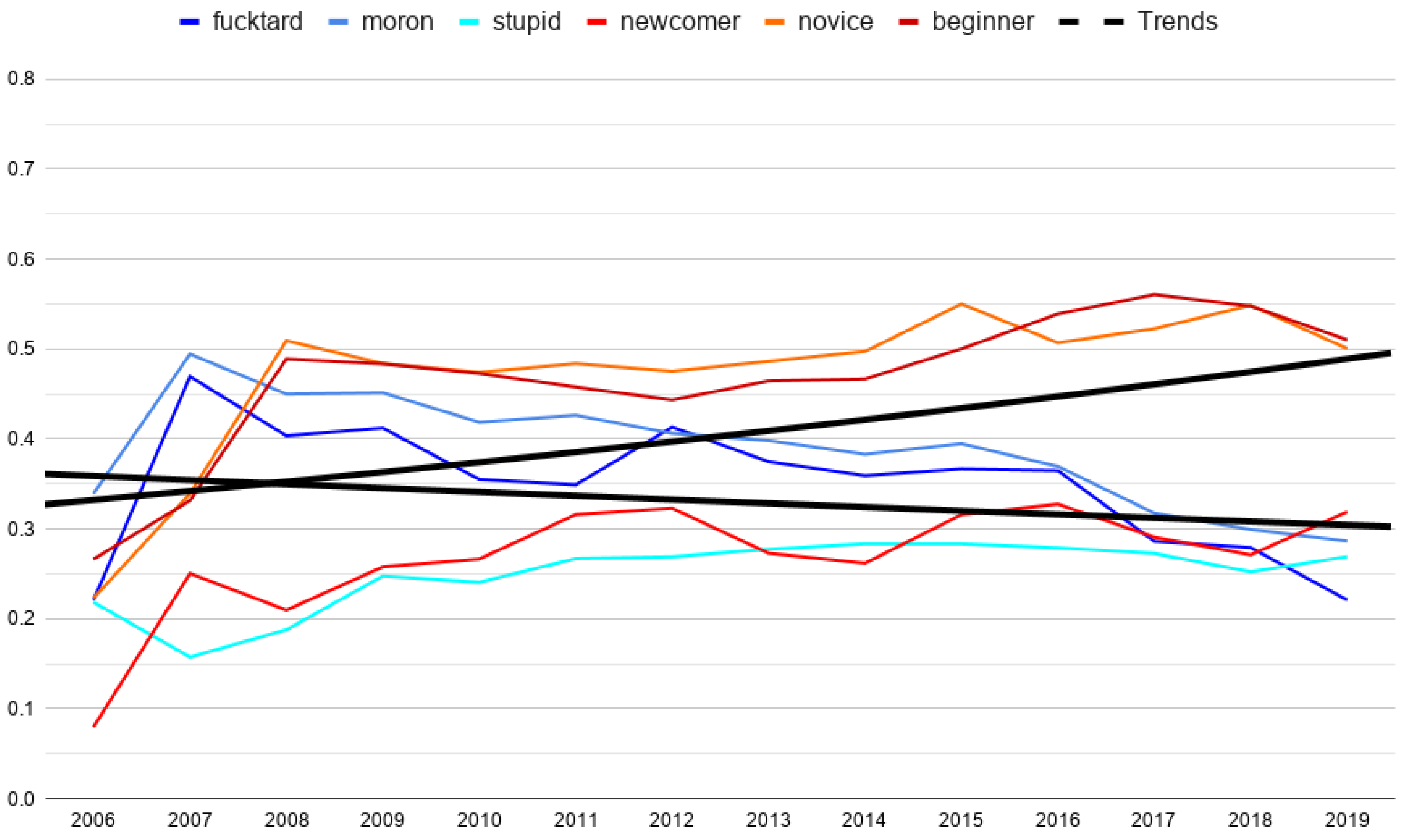

Figure 8. Similarity scores for *noob* and the exponential trends of the averages of the senses. The increasing trend represents the neutral sense and the decreasing trend represents the negative sense.

The average score of negative contexts (*moron, fucktard, stupid)* begins with 0.4 until, though decreases to 0.2 by the end of the timeline. This change is more stable than the average score for the neutral contexts: the similarity between *noob* and *newcomer, novice* and *beginner* begins at 0.2 and steadily rises to 0.5 by the end of the timeline.

The analysis shows that *noob* may have seen *elevation* to some degree. Both the similarity scores and the vector representation seem to evidence the change of *noob* previously occurring within more negative contexts, and more recently occurring within neutral contexts. Nevertheless, while there is an observable change in these scores, the overall increase or decrease for the senses is approximately 0.2. With this, it remains to be seen if this change will continue and the senses become more apart.

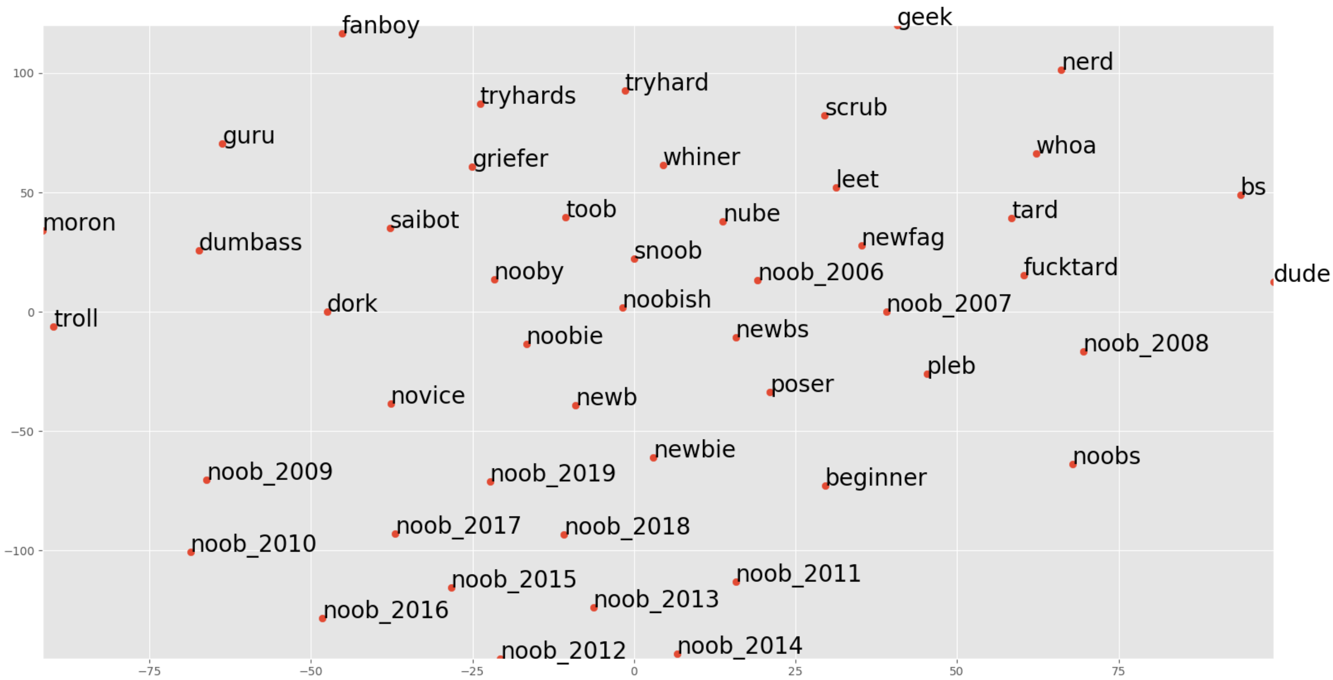

Figure 9. Two-dimensional representation of the vector space of the word *noob.*

According to the vector space representation, the first three years are especially close to the negative senses, and there is later a rather notable leap towards the neutral senses. Afterwards, the change is relatively inconsistent—as in it is difficult to draw a linear chronological path between each of the *noob* vectors. On the other hand, the *noob* vectors for each year seem to revolve closely around the neutral senses. The geometrical distribution of the words is not clear and the representation is populated with variants of *noob.* Nevertheless, there is still an observable change of the most recent positions of *noob* being closer to *novice, beginner* and other variants of *noob*.

4.4 Lit

There seems to be change with the word *lit* as well. In fact, the change happens during the last half of the timeline. Specifically, *lit* seems to gain a new sense which is closely synonymous with *good* or *awesome*. With this, the most similar words to *lit* are related to fire and light (*illuminated, lighted, candles, burning,* etc.), though after 2015 the word *dope* emerges as one of the most similar words, and, in fact, is the third most similar word to *lit* in 2019.

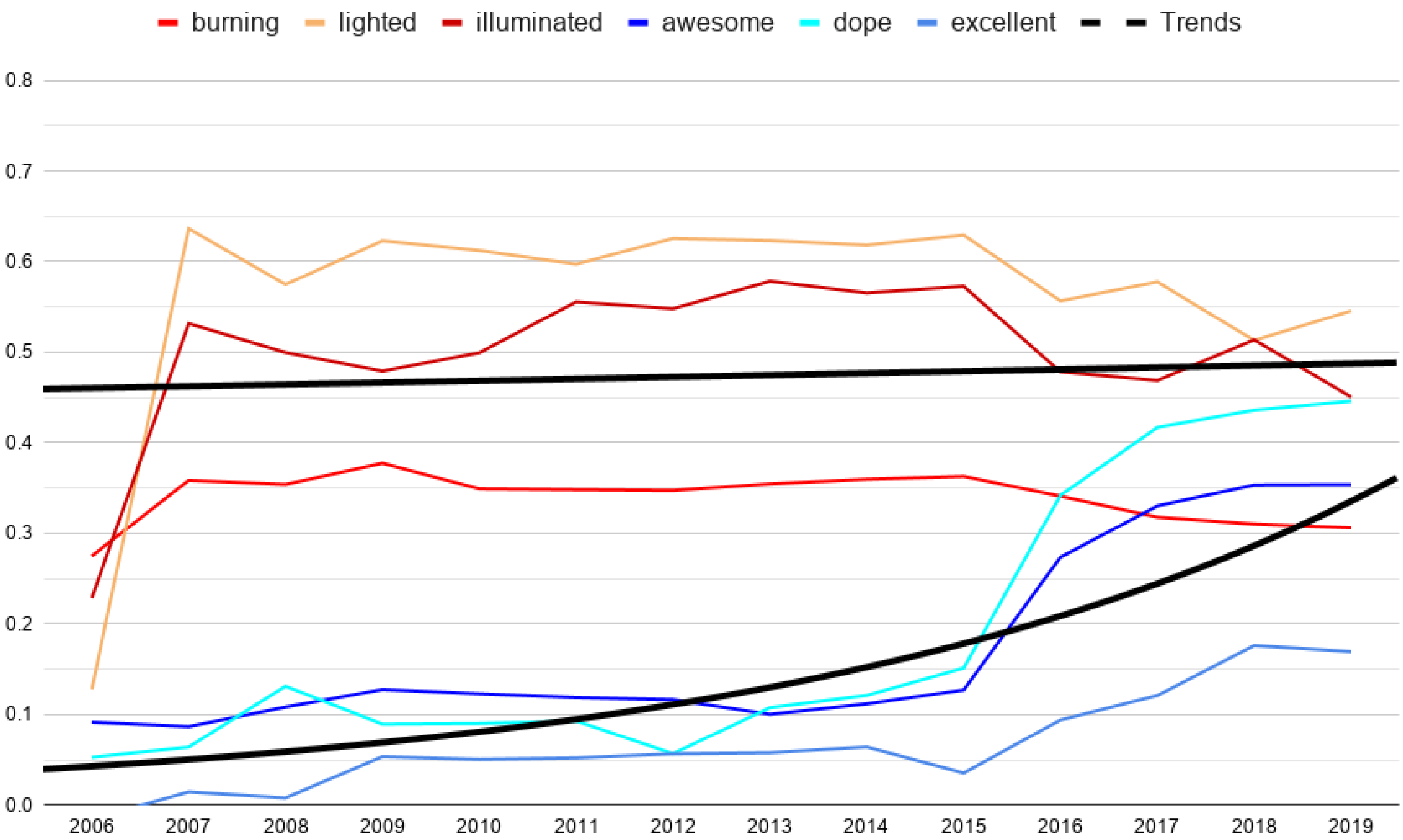

Figure 10. Similarity scores for *lit* and the exponential trends of the averages of the senses. The increasing trend represents the sense of "really great" and the decreasing trend represents the sense of "lighted".

This is a relatively recent change since the new sense seems to emerge in 2015 and the similarity scores between *lit* and *dope* are less than 0.1 before the year 2015, yet increases to 0.4 by the end of the timeline. With this, the similarity scores for the original sense are rather stable, decreasing to 0.4 from 0.5. In this sense, it seems that both of the discussed senses of *lit* are being used approximately equally often.

Lit undergoes similar change as *fam,* where the word gains a new sense which has its origins from slang or a language variety (AAVE). With this, Merriam-Webster describes the contemporary meaning of *lit* as “extremely great.” Similarly to *fam,* the use of *lit* originates from AAVE. Regardless, *lit* could be argued to have undergone *elevation* to some extent*,* since the new sense is very positive. Nevertheless, this new sense does not replace the original senses and the change is not perhaps a traditional type of semantic change, but an appropriation of AAVE, and hence the term is used by non-AAVE speakers. Because of this, categorizing *lit* in this way may not be appropriate, as the new observed use of *lit* has existed before, yet it was not appropriated into a more standard variety of English.

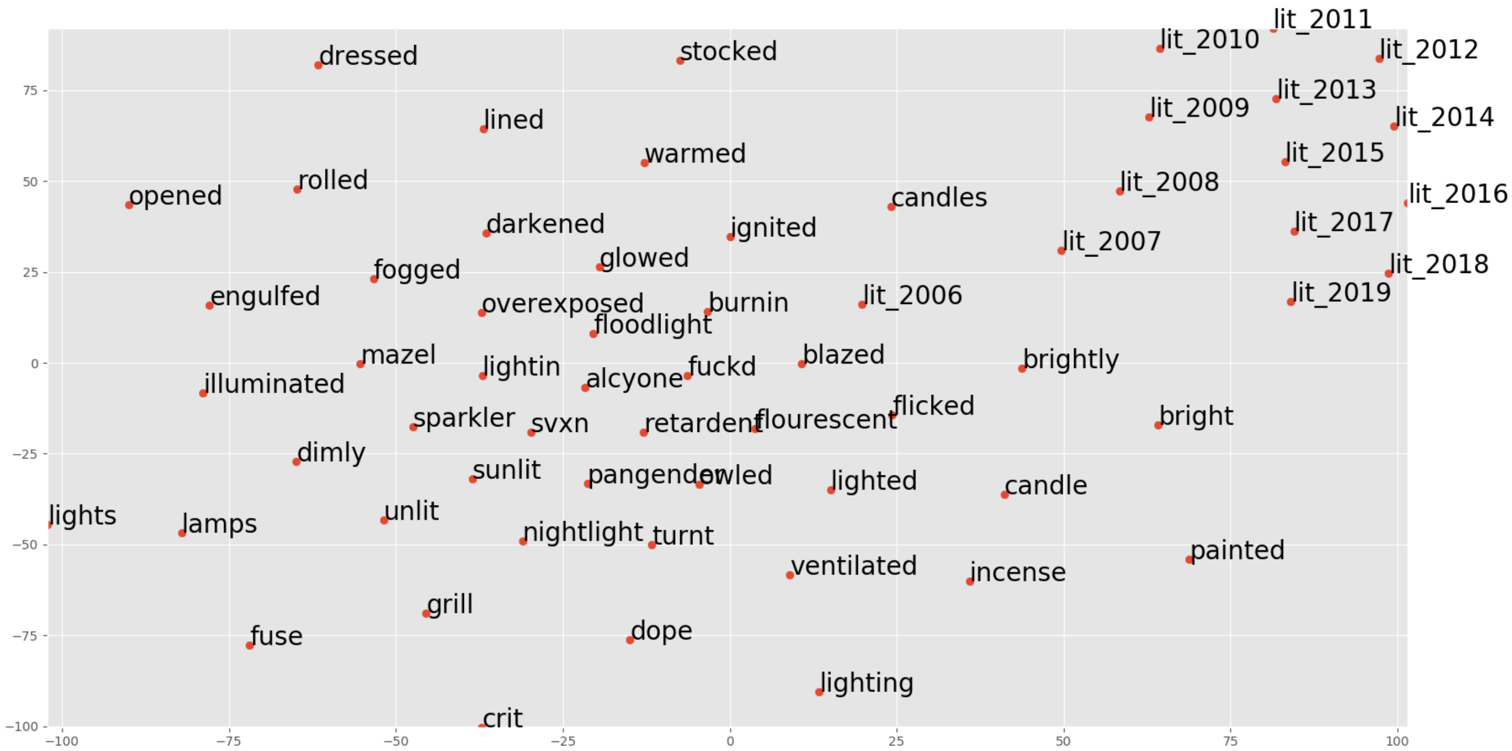

Figure 11. Two-dimensional representation of the vector space of the word *lit.*

The change is rather consistent according to the vector space representation, yet due to the nature of the 300-dimensional vectors being projected into two dimensions, the similarity with the word *dope* is very difficult to visualize. Because of this, *dope* is the only word representing the new sense, while most of the other neighboring words are related to fire and light. In this way, the vector representation is mostly populated with words of the original context, and hence the change is difficult to observe from vector space alone.

4.5 Sick

With *sick*, we may observe that the often discussed sense that is related to *cool* (Mitra, et al. 2014) is already present in the corpus throughout the timeline. In addition, the similarity for the traditional sense of *sick*—being ill or unwell—is quite high throughout the timeframe, and does not seem to be replaced by the *cool* sense.

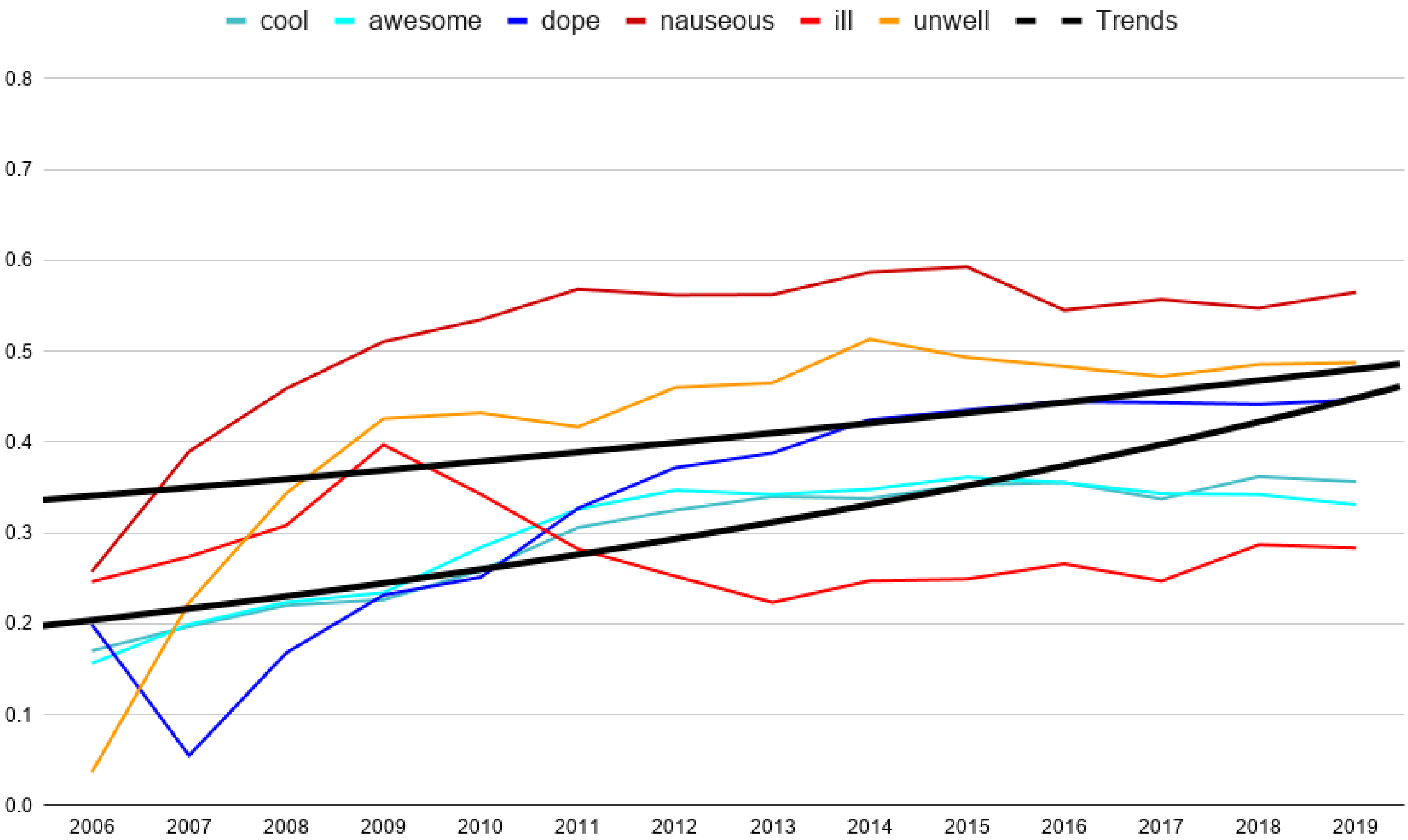

Figure 12. Similarity scores for *sick* and the exponential trends of the averages of the senses. The upper increasing trend above represents the sense of “ill” and the increasing trend below represents the sense of “cool.”

The similarity between *sick* and the senses representing the newer sense increase to 0.4 from 0.1. For the original sense, the similarity scores do not change much, and each similarity score stays within its own decimal score. For example, the similarity score between *sick* and *nauseous* stays within 0.5. In fact, if the years between 2006 and 2009 were omitted from the similarity scores, there would be no increasing or decreasing trend for the original score. Nevertheless, the trend of the new score is slowly increasing. Because of this, it is more likely that the word embeddings at the beginning of the timeline are not accurate, and there may be a lack of data causing the similarity between *sick* and *unwell* to be relatively low at the beginning of the timeframe.

With this, the word *sick* has been affected by *elevation,* in which the original negative sense of being ill or unwell. While this change does not specifically emerge within the examined timeframe, nor do the words *cool, dope* or *awesome* appear within the projected vector space, the similarity scores still give evidence for the change that has already occurred, as the trend of the new sense is still increasing.

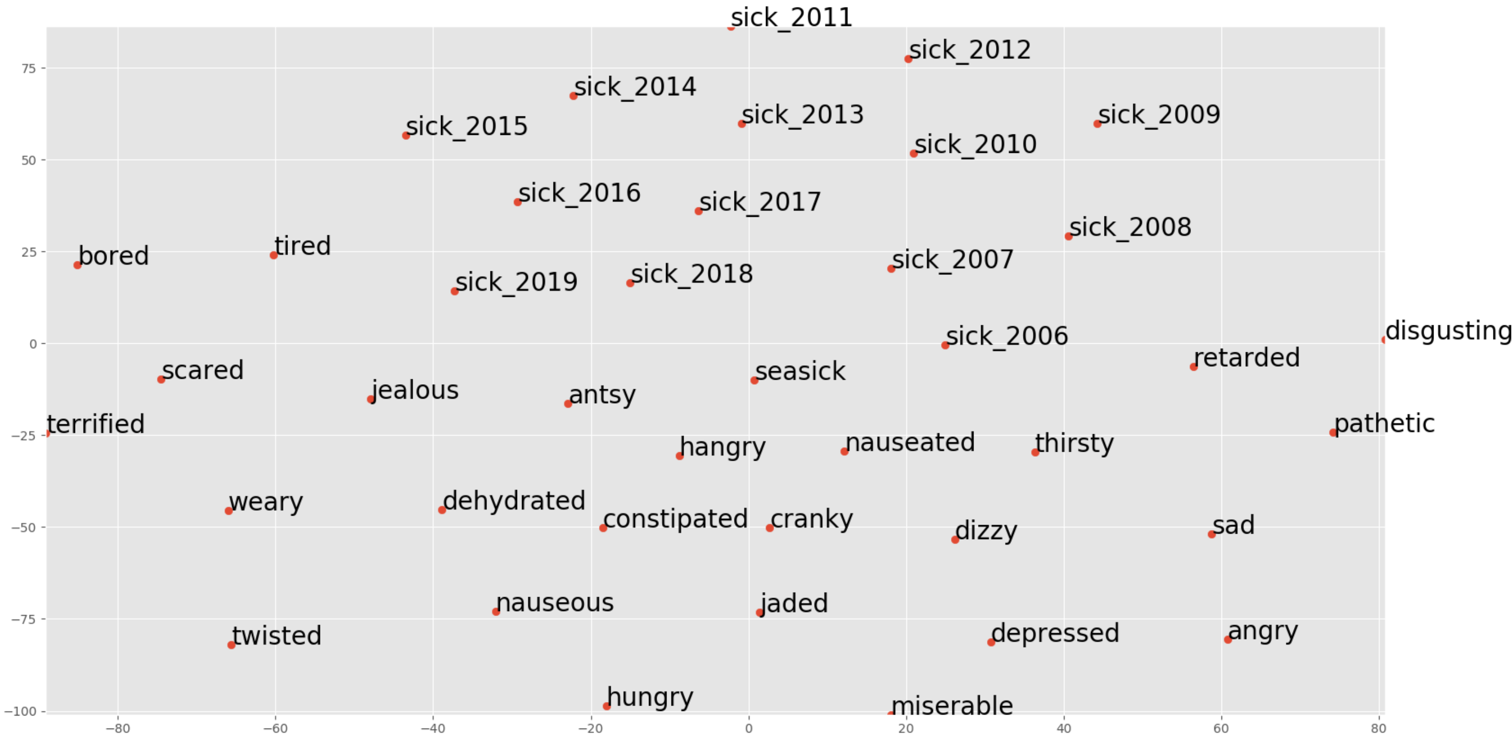

Figure 13. Two-dimensional representation of the vector space of the word *sick.*

The vector space representation and the most similar words to *sick* in most recent years, nevertheless, reveal some new information for analysis. In fact, the word *sick* seems to be moving relatively consistently towards *bored, tired* and *jealous.* In fact, the similarity scores between *sick-bored* and *sick-tired* grow consistently from approximately 0.3 to 0.6 by the end of the timeframe. This could be evidence for a context where *sick* is used followingly: “I’m so sick of __.”. With this, a speaker would be expressing their dissatisfaction with something. In this sense, the aforementioned words would work in a similar context, hence this sense of *sick* may be emerging as a new one. In this case, this semantic change could be labeled as *hyperbole*, where the original meaning is weaker and whereas the new, stronger sense causes the word to be used in an exaggerated way.

4.6 Toxic

The most similar words to *toxic* throughout 2006 to 2012 include *poisonous, carcinogenic, radioactive, hazardous, dioxins, corrosive.* Then, the list of most similar words changes into *hateful, immature, harmful, disrespectful, misogynistic, judgemental* and *abusive.* With this, we can see a very clear change of the most common sense from the sense of toxicity in the context of “poisonous substances” to the context of e.g. “harmful person or relationship.”

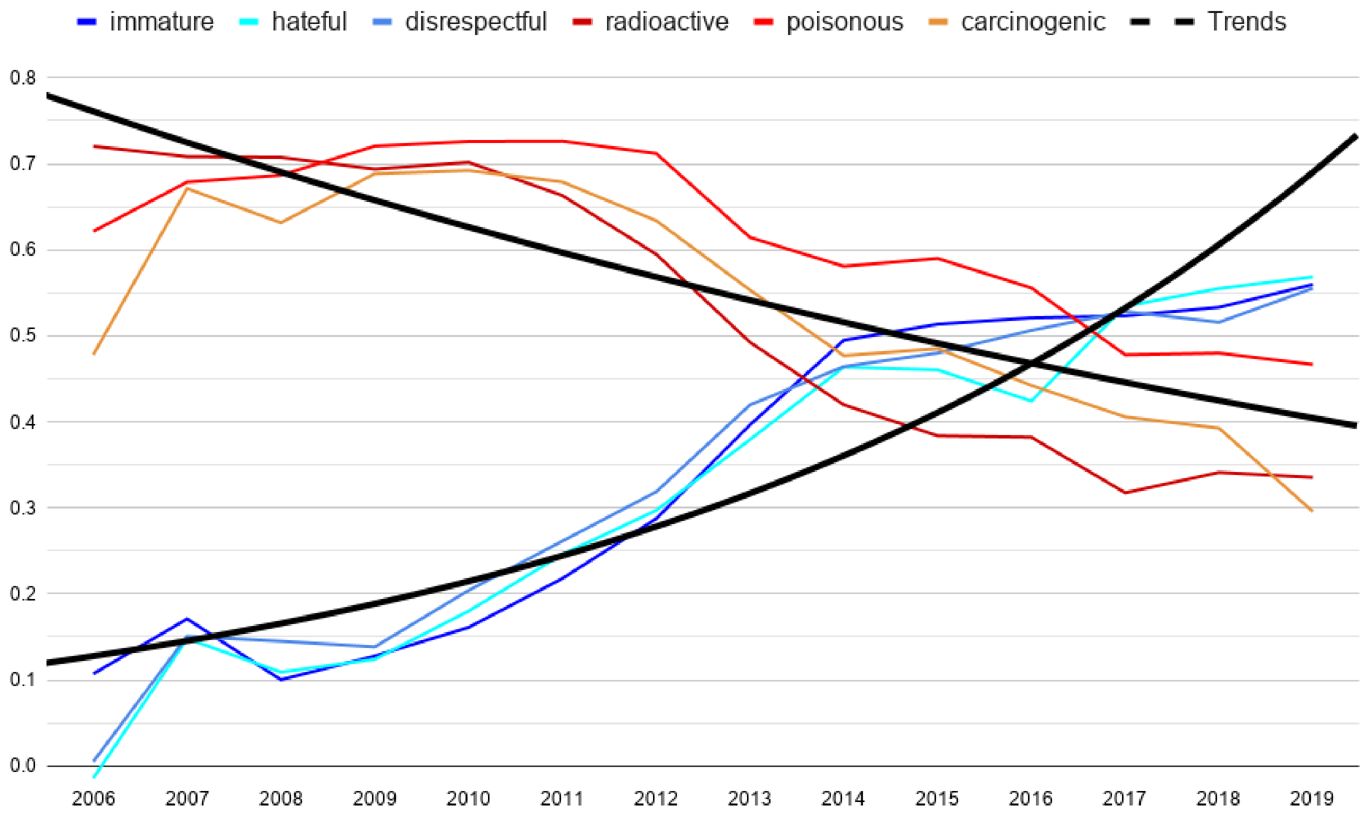

Figure 14. Similarity scores for *toxic* and the exponential trends of the averages of the senses. The increasing trend represents the sense of “disrespectful behavior” and the increasing trend below represents the sense of “poisonous.”

It can be seen clearly that the new sense of *toxic* is quite low in the beginning. In fact, the similarity between *toxic* and *hateful* remains less than 0.2 before 2011. On the other hand, the similarity scores for *toxic* and *poisonous, radioactive* and carcinogenic average to 0.7 for the same timeframe. As the scores move towards the end of the timeline, the scores are more similar, yet the new sense is still more common: an average of 0.5 and above for the new sense and an average of 0.4 for the original sense.

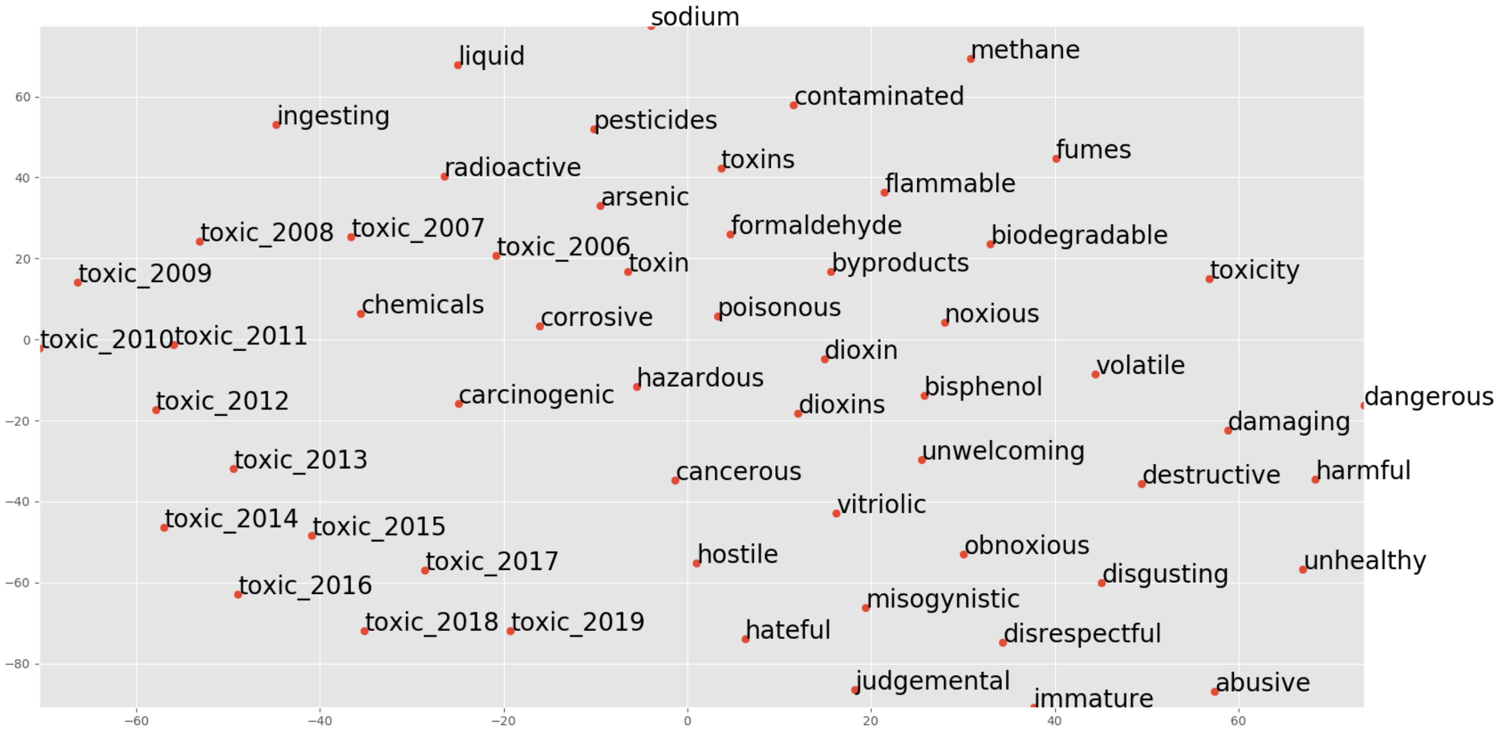

Figure 15. Two-dimensional representation of the vector space of the word toxic.

The change can be visualized relatively well on the vector space representation. On the top left of the representation, there is a grouping of similar words related to the traditional sense of chemical toxicity, such as *radioactive, arsenic, toxin, chemicals, corrosive*, and on the bottom right there is a grouping of neighboring words related to the emerging sense, such as *obnoxious, disrespectful, immature, judgemental, abusive.* Moreover, the word *toxic* seems to clearly move consistently and gradually towards the group of neighboring words representing this new sense of toxic people or relationships.

Toxic has seen a *metaphorical* change. In this way, a toxic relationship between people may not be exactly poisonous in a chemical sense, but it could be expressed as such metaphorically. There could be some arguments for the category of *degeneration* occurring with *toxic*, since this emerging sense describes a relationship that is most of the time undesirable and seen as negative between people, whereas poisonous substances can be seen as neutral (e.g. insect repellents or the self-defense of plants).

5 Discussion

By inspecting the way senses in all of the analyzed words change distributively, we have gained some evidence for detecting semantic change using word embeddings. This distributive framework lends itself appropriately to noticing changes in word collocational patterns, with both of the methodologies employed in this paper seem to suggest some changes in word meaning. Firstly, in cases of *swipe, lit, fam* and *toxic*, the change in the similarity of a context is relatively consistent and gradually changing. In this sense, the chronological paths for the target words in the projected two-dimensional representations can be seen quite clearly. For instance, the vectors for *toxic* from 2006 to 2019 seem to follow a definite path and the pattern is not unpredictable in this sense. Moreover, this consistent change is aligned with the comparison of word similarities, in that a consistent change within the vector representation can also in some cases be seen in the change of word similarity. Specifically, as *toxic* moves consistently further away from the chemical senses on the representation, the word similarity also decreases gradually.

Bloomfield introduced the types of semantic change within the context of cognates and language families. The effect of applying these categories within timeframe in a different context, however, can be observed in terms of how suitable they are in classifying these changes. In fact, not all the categories seem to be explicitly applicable in this paper. Generally, these categories contain the presupposition that the original sense is mostly replaced by the new sense. This is exemplified by the previously mentioned *mete* “food” changing into the word *meat.* In this way, the *narrowing/broadening* category may be problematic if it is applied in a different context, yet in a similar manner, as Bloomfield. Specifically, this is due to a word form unlikely undergoing a change in which the meaning is changed fully. In other words, for *narrowing/*broadening to be applicable, the original sense (e.g. “food”) may have to be completely replaced by the new sense (“meat).

The types of semantic change that seem to be the most applicable are *elevation/degeneration* and *metaphor*. This is due to word forms being capable of acquiring new senses without the new senses completely replacing the original ones. For instance, with *swipe* and *toxic*, the emerging senses can be observed quite clearly, yet the similarity scores of the original senses imply that the original sense is still in use. As for *elevation/degeneration*, the original sense may change into acquiring a more positive or negative connotation, as evidenced by, for example, there being a different amount of slurs and offensive words as the closest neighboring words. The word *noob* is an example of this, in which the most similar words have changed from slurs to neutral expressions. With this, after recontextualizing *elevation/degeneration* and *metaphor* for shorter periods, the categories seem to be sufficient to be operationalized in classifying semantic change.

Because of the nature of the selected words, the change in some of the words can be argued to occurr due to non-linguistic causes. Namely, *swipe* contains a new sense due to technological innovation, and *lit* and *fam* adopt the senses from a language variety. In this way, some of the categories are not equally applicable in categorizing short term semantic change as they are when often employed to categorize historical, linguistic change. Moreover, the recent senses for *lit* and *fam* were introduced from AAVE (Merriam-Webster, 2017), hence the meanings for the word forms have existed prior to the analysed data. Nevertheless, the meanings did not exist in the more standard variety of English analyzed in this paper. In this instance, by observing the introduction of a new sense from another language variety one may be required to employ different categories than ones introduced by Bloomfield. On the contrary, it could be seen that *fam* was subjected to *broadening* in the same way historical linguistics claims cognates to change between languages. In this sense, the *narrowing/broadening* category must be recontextualized in the way that we are also sensitive to how these changes and new senses emerge.

It seems that the distributional balance of the word senses is a key factor in how appropriate the types of semantic change are in categorizing change. In the same way *narrowing/broadening* is difficult to observe in short term change, *substitution* seems to contain a similar problem. In fact, no words were observed to be subject to this type. While *substitution* occurs in the context of cultural change—which may seem appropriate for the context of the selected words—the original senses are still not completely replaced by the new ones. Arguably, there could be some kind of technological innovation that would cause a full change in word meaning in a decade, but this type of word was not present in the paper. With *swipe,* this could be the case in the future if the sense of user interaction completely replaced *swiping* in the sense of a motion or e.g. a sword attack, but similarly to the type of *narrowing/broadening*, this is unlikely to happen within a decade.

In terms of the models, some large distributional changes and shifts in word meaning can be seen between 2006 and 2009. Moreover, there is less data for these years, and the frequency of the target words increases markedly after 2009. Most importantly, the parameters for all the models are still the same, due to the vector spaces having to be aligned. In this way, while there is much less data in 2006, the dimensionality and the window size of the model for that year are the same than the models between 2010 and 2019. This may cause the fact that the vector spaces are not directly compareable, since dimensionality with large data will produce different vectors than with a small amount of data. Because of this, the results of this paper seem to suggest that future models based on TWEC and word2vec should contain equal amount of data distributed between the time intervals.

Frequency may be a factor to be examined in future studies. In fact, the changes in word similarities and sense distributions seem to be highly correlated with the frequencies of the target words. For example, while the most common sense of *swipe* changed from an association of paying by a credit or debit card to a context of user interface interaction, the frequency of *swipe* increased. In this way, the frequency was highly correlated all the individual word similarity scores, as well as the average similarity score. In Pearson's correlation coefficient, the statistical relationship was 0.7 to 0.9 for the increasing senses and -0.7 to -0.9 with the decreasing senses and increasing frequency. In this way, the development of technology may cause a cultural change, which in turn causes people to use a word form in a different way. In particular, the word is used in more varied contexts, and thus *swipe* gains additional polysemy through *metaphorical* change. Generally, this would indicate that language change causes polysemy, while this is quite opposite of what Hamilton, et al. argue with their results (polysemy leads to language change).

Though this paper has observed regularities and consistent changes in some of the target words, none of the changes or variables have necessarily been statistically validated in terms of the rate of change. For more meaningful results, future research could benefit from a methodology that results in more statistical power. Moreover, the method of choosing three words, which occurr exclusively in the same specific context as a target word, can be seen as somewhat arbitrary, and there may be a statistically significant and accurate way of representing specific contexts of a word. Nevertheless, the current method has produced results and is more sophisticated than merely looking at single word similarity pair scores. These single scores may not be representative of the whole context or sense of the inspected target word.

6 Conclusion

This paper shows the following possible types of semantic change for the selected words: (1) *metaphorical: swipe, toxic, (2) elevation: noob, lit, sick* and (3) *broadening: fam*. There could also be a case argued for *toxic* undergoing *degeneration.* With these changes, there is a varying degree to how appropriate it is to attempt to categorize these words and apply the categories themselves. This is directly connected to the context in which Bloomfield first employed these types of semantic change. Whereas Bloomfield discusses e.g. *narrowing/broadening* as a historical linguistic change, generally, the distributional changes in the senses examined in this paper may not have the time frame to largely replace another sense. Because of this, the current paper has examined the more nuanced balance of these senses. Specifically, it may be so that e.g. *narrowing*/*broadening* are mostly linguistic types of changes which can not be applied without recontextualizing the categories to fit a shorter time frame. On the other hand, by using word embeddings, it is possible to observe new contexts in which a word is used, hence implying new word senses emerging. Thus, categories such as *metaphorical* change and *elevation/degeneration* lend themselves appropriately to classifying semantic change in this way, since it is plausible for words to gain new senses within a short period.

The methodology employed in the current paper may be legitimate in terms of providing evidence for semantic change that is either already known or that a linguist might have intuitions about before researching a particular word. In this way, this does not enable someone to automatically detect unknown changes. Nevertheless, this paper hopefully shows that a linguist may test and examine the degree to which a hypothesized semantic change has occurred. While this is the case, strictly comparing word pair similarity scores and the averaging of these scores to form a specific context has not been to the author's knowledge been done before. In this sense, a statistical method to validate the data and the change in the senses may be appropriate to make similar research provide statistically more meaningful results.

While the results in this paper suggest that it is possible to observe semantic change by using word embeddings, research employing word embeddings to research semantic change should be taken further from merely detecting semantic change. If this is done by classifying semantic change, future work should contextualize the categories for the type of research that is being done with the word embeddings. As stated by Kutuzov, et al (2018, p. 10), there has been a tendency within the field to detect semantic change, but not continue the analysis any further (also argued by Boleda 2020.) While this has been addressed to some degree by Mitra et al. (2014), this paper agrees with the notion by Kutuzov, et al. that "much more work is certainly required to empirically test classification schemes" (p. 10.) While the current paper has pursued this by evaluating some of the schemes, much of the other theoretical work needs to be examined as distributional methods continue to increase as a methodology.

7 References

Bloomfield, L., 1933. Language. 1933. *New York: Holt.*

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. and Blackburn, J., 2020. The pushshift reddit dataset.

Boleda, G., 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics.*

Boleda, G. and Herbelot, A., 2016. Formal distributional semantics: Introduction to the special issue. *Computational Linguistics, 42*(4), pp.619-635.

Barthel, M., Stocking, G., Holcomb, J. and Mitchell, A., 2016. Reddit news users more likely to be male, young and digital in their news preferences. *Pew Research Center.*

Di Carlo, V., Bianchi, F. and Palmonari, M., 2019, July. Training temporal word embeddings with a compass. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 6326-6334).

Ferrone, L., Zanzotto, F.M., 2020. Symbolic, Distributed, and Distributional Representations for Natural Language Processing in the Era of Deep Learning: A Survey. Front. Robot. AI 6, 153. https://doi.org/10.3389/frobt.2019.00153

Firth, J.R., 1957. A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*.

Gulordava, K. and Baroni, M., 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* (pp. 67-71).

Hamilton, W.L., Leskovec, J., Jurafsky, D., 2018. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.

Hamilton, W.L., Leskovec, J., Jurafsky, D., 2016. Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change.

Harris, Z.S., 1954. Distributional structure. *Word*, *10*(2-3), pp.146-162.

Hilpert, M., 2008. Germanic future constructions: A usage-based approach to language change (Vol. 7). John Benjamins Publishing.

Merriam-Webster, 2020, “fam”, Accessed May 14th, 2020, https://www.merriam-webster.com/dictionary/fam

Merriam-Webster, 2017 It’s lit – An old slang word made new again, Accessed May 14th 2020, https://www.merriam-webster.com/words-at-play/lit-meaning-origin

Kutuzov, A., Øvrelid, L., Szymanski, T., Velldal, E., 2018. Diachronic word embeddings and semantic shifts: a survey.

Smith, S. L.; Turban, D. H.; Hamblin, S.; and Hammerla, N. Y. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax.

Stern, G. 1931. Meaning and change of meaning; with special reference to the English language. Wettergren & Kerbers.

Github Repository for the paper: https://github.com/Teemursu/reddit_neologism_semantic_change

TWEC: https://github.com/valedica/twec