

**LONGITUDINAL COMPARISON OF WORD ASSOCIATIONS IN
SHALLOW WORD EMBEDDINGS**

by

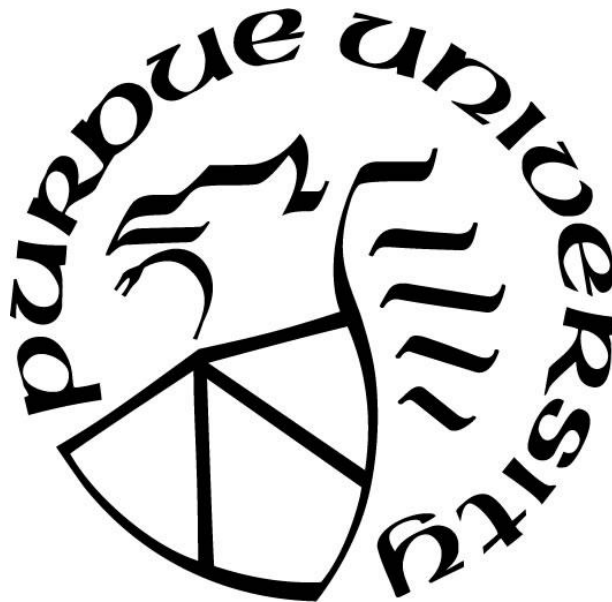
Geetanjali Bihani

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science



Department of Computer and Information Technology

West Lafayette, Indiana

May 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Julia T. Rayz, Chair

Department of Computer and Information Technology

Dr. Baijian Yang

Department of Computer and Information Technology

Dr. Ida B. Ngambeki

Department of Computer and Information Technology

Dr. J. Eric Dietz

Department of Computer and Information Technology

Dr. Lisa Bosman

Department of Technology and Leadership & Innovation

Approved by:

Dr. Eric T. Matson

*To my parents,
whose prayers, love and encouragement support my journey and inspire me to work towards
achieving my goals*

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my graduate advisor, Professor Julia T. Rayz, who provided expert guidance and encouragement throughout my research.

I would also like to thank my committee members, Professor Baijian Yang, Professor Ida B. Ngambeki, Professor J. Eric Dietz and Professor Lisa Bosman, for their valuable suggestions and support.

I would also like to thank my lab members at the AKRaNLU Lab, who provided useful feedback during the ideation of this work.

Finally, I would like to thank my friends and family, for being constant supporters and motivators as I go through this learning experience.

TABLE OF CONTENTS

LIST OF TABLES	7
LIST OF FIGURES	8
LIST OF ABBREVIATIONS	9
GLOSSARY	10
ABSTRACT	11
INTRODUCTION	12
Research Questions and Objective.....	13
Assumptions.....	14
REVIEW OF LITERATURE	15
Introduction.....	15
Word2vec.....	16
Glove.....	16
fastText	17
Biases in word embeddings	17
Embedding level biases	17
Corpus level biases	18
Bias Identification Methods	18
Word Embedding Association Tests	18
Analogy based identification	19
Bias sub-space identification	19
Bias evaluation datasets	20
Crowdsourced bias identification	20
Bias Resolution Methods	21
Debiasing as a post-processing task	21
Debiasing as a training task	22
Corpus augmentation	22
Gaps in Previous Research.....	23
METHODOLOGY	25
Text corpus: Corpus of Historical American English.....	25

Labeling neutral and attribute words	26
Part of Speech Tagging.....	27
Training word embeddings	28
Creating neutral and attribute word lists/clusters.....	29
Measuring word associations	31
Quantitative Measures	31
Qualitative Measures	33
Evaluation Framework.....	33
RESULTS	36
Evaluation I: Variation in word associations over varying word embedding models	36
Evaluation II: Variation in word associations over varying text corpora	45
Future Work	46
CONCLUSION.....	48
REFERENCES	50
APPENDIX A. TOP 10 ELEMENTS FOR EACH CLUSTER	53
APPENDIX B. JACCARD SIMILARITY BETWEEN WORD EMBEDDING CLUSTERS ...	98

LIST OF TABLES

Table 1 COHA corpus word distribution.....	26
Table 2 Model Hyperparameters	28
Table 3 Evaluation framework.....	33
Table 4 Dunn’s Index over varying word embedding models and fixed text corpus	34
Table 5 Dunn’s Index over varying text corpora and fixed word embedding model	34
Table 6 Dunn's Index for word embedding model clusters, $N = 20$	36
Table 7 Average Jaccard Similarity between word embedding models (1810s – 2000s)	40
Table 8 Rare words clustered together (Word2vec (CBOW), 1910s, Cluster 3)	41
Table 9 Rare words clustered together (Word2vec (Skip-gram), 1940s, Cluster 3).....	42
Table 10 Person names clustered together (Word2vec embeddings)	43
Table 11 Place/Location names clustered together (Word2vec embeddings)	43
Table 12 Group names clustered together (Word2vec embeddings).....	43
Table 13 Syntactically similar words clustered (fastText)	43
Table 14 Variation in Dunn's Index over decades	45

LIST OF FIGURES

Figure 1 Plutchik's wheel of emotions (Maupomé & Isyutina, 2013)	30
Figure 2 Plot of Dunn's Index over 20 decades (1810s - 2000s)	37
Figure 3 Distribution of words across embedding clusters.....	39
Figure 4 Variation in Dunn's Index across word embeddings (1810s -2000s).....	46

LIST OF ABBREVIATIONS

CBOW – Common Bag of Words

COHA – Corpus of Historical American English

DI – Dunn’s Index

ELMo – Embeddings from Language Models

GloVe – Global Vectors

NLP – Natural Language Processing

POS – Part-of-Speech

WEAT – Word Embedding Association Test

GLOSSARY

Clustering – the task of grouping a set of objects in such a way that objects in the same group, which is called a cluster, are more similar to each other than to those in other groups

Natural language processing – a component of artificial intelligence that aims at having the computers to analyze, understand and derive meanings from human languages

Part-of-Speech – a category of lexical items which have similar grammatical properties

Semantic similarity – distance metric based on the likeness of the meaning of a set of terms or documents

Syntactic similarity – distance metric based on the likeness of the string format of a set of terms or documents

ABSTRACT

Word embeddings are utilized in various natural language processing tasks. Although effective in helping computers learn linguistic patterns employed in natural language, word embeddings also tend to learn unwanted word associations. This affects the performance of NLP tasks, as unwanted word associations propagate and amplify biases. Current word association evaluation methods for word embeddings do not account for changes in word embedding models and training corpora, when creating the rubric for word association evaluation. Current literature also lacks a consistent training and evaluation protocol for comparison of word associations across varying word embedding models and varying training corpora. In order to address this gap in prior literature, this research aims to evaluate different types of word associations, not limited to gender, racial or religious attributes, incorporating and evaluating the diachronic and variable nature of words over text data collected over a period of 200 years. This thesis introduces a framework to track changes in word associations between neutral words (proper nouns) and attributes (adjectives), across different word embedding models, over a temporal dimension, by evaluating clustering tendencies between neutral words (proper nouns) and attributive words (adjectives) over five different word embedding frameworks: Word2vec (CBOW), Word2vec (Skip-gram), GloVe, fastText (CBOW) and fastText (Skip-gram) and 20 decades of text data from 1810s to 2000s. Finally, various cluster level and corpus level measurements will be compared across aforementioned word embedding frameworks, to find how word associations evolve with changes in the embedding model and the training corpus.

INTRODUCTION

Data are the building blocks of knowledge in artificially intelligent systems. One of the richest sources of such data is natural language, the foundation of human knowledge and the conveyor of human thoughts and beliefs. However, in order for machines to learn from human knowledge, it is essential for them to learn the semantic and syntactic structures employed within the natural languages conveying that knowledge.

Computers cannot decipher natural language data in its textual format. In order to convey the information hidden within such data and bridge the gap between natural language and computer understanding, text data need to be converted into word representations that can be understood by computers. To create such word representations, word embeddings are increasingly being employed. Word embeddings provide an efficient way to pack dense information available within textual data into vectors. By considering the distributional hypothesis that words used in similar contexts tend to have similar meanings (Harris, 1954) word embeddings capture semantic and syntactic regularities in language (Mikolov, Yih, & Zweig, 2013).

Unfortunately, these embeddings are also susceptible to the biases present within natural language data (Caliskan, Bryson, & Narayanan, 2017). Biases derived from societal prejudices are reflected within natural language. These biases and stereotypes, when utilized as an input to any embedding algorithm, affect the fairness of the respective embedding framework. If the natural language data used to train a model does not correctly represent the reality, it leads to an unfair and inaccurate model being built (Gordon & Van Durme, 2013). These acquired biases affect downstream NLP tasks such as classification (Prost, Thain, & Bolukbasi, 2019), resume sorting system (Bolukbasi, Kai-Wei, Zou, Saligrama, & Kalai, 2016) and loan eligibility and crime recidivism predictions (Chouldechova, 2017), to name a few. Thus, a major concern while creating word embeddings are the plethora of unwanted and inappropriate biases that are encapsulated during the embedding creation process.

Prior literature has proposed different methods to identify and reduce such biases. Some methods utilize analogy based word association testing to identify biases within text (Caliskan et al., 2017), while others base their methods on gender direction based bias identification and debiasing (Bolukbasi et al., 2016). Since each method uses a different text corpus to conduct

their analyses, it is difficult to compare their processes. Moreover, each method uses a different set of metrics to identify bias, leading to a disparate set of bias analysis approaches. Another factor that heavily impacts bias in word embeddings is the word embedding creation framework, which affects the information contained in each embedding. This has led to a lot of gaps in the research and demands for a consistent protocol which can identify the pros and cons of various bias identification and debiasing approaches by comparing and contrasting across a consistent text corpus.

Research Questions and Objective

This thesis aims to identify how properties of a text corpus and word embedding models inform word associations generated between different words. The main focus will be identifying the relationship between corpus level parameters and proposed word associations evaluation metrics, which can identify how word embedding models capture word associations, which can help inform frameworks for robust debiasing processes. By identifying the diachronic nature of word embeddings, context and usage driven analysis of word associations can be conducted. Since word associations within word embeddings are a direct by-product of the context in which the words are being used in, it also gives us a view of how word-context evolution affects evolution of word associations. This research compares word associations captured across five different widely used shallow word embedding frameworks. It aims to answer two research questions:

Q.1. How do word associations differ across varying word embedding frameworks, while keeping a consistent training corpus?

Q.2. How do word associations differ across the same word embedding framework, while varying training corpora?

Overall, this thesis aims to contribute towards the development of trustworthy and fair artificially intelligent systems by advancing the processes of identifying the difference between biased and unbiased word associations within word embeddings, a primary input component in such systems.

Assumptions

Certain assumptions have been made while conducting this research. These have been listed below:

1. The StanfordNLP Pipeline POS tagger (v0.2.0) provides acceptable levels of labeling accuracy, when labeling entities and parts of speech on given text inputs.
2. The COHA text corpus for a given decade is representative of the natural language employed during that particular time period.

REVIEW OF LITERATURE

In order to create a holistic picture of the treatment of biases in word embedding representations, a comprehensive review of literature is required. This review aims to identify and critique the effectiveness of recently proposed word embedding bias identification and bias resolution methods.

The scope of this literature review is limited to the prior literature on the analysis of different types of biases in English language word embeddings. Three specific word embedding frameworks, namely Word2vec, GloVe, and fastText, are described. This has been done because these are widely popularized word embedding models and are heavily used in building applications and solutions to solve various NLP problems. During the literature collection process, it was observed that a large body of prior literature has focuses on bias analysis for Word2vec, GloVe and/or debiased versions of these primary word embedding frameworks. Thus, this literature review aims to not only conduct a comparison of bias analysis methods by addressing the benefits and gaps of those methods, but also identify robust bias evaluation and debiasing methods which can be extended to the latest word embedding frameworks, as future work.

Introduction

Word embeddings are distributed word representations which allow for each word in a given text vocabulary to be mapped to a vector in a low-dimensional space. The initial concept of word embeddings was proposed in 2003, which allowed learning of a distributed representation of each word as well as the probability function of word sequences (Bengio, Ducharme, Vincent, & Jauvin, 2003). The authors also claimed, “the work presented here opens the door to improvements in statistical language models brought by replacing “tables of conditional probabilities” by more compact and smoother representations” (Bengio et al., 2003). By 2008, the process of creating word embeddings was improved by combining a neural network architecture to build generalizable word embeddings (Collobert & Weston, 2008). But it was not until 2013, when the concept of word embeddings reached new heights by the popular Word2vec model (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Since then, multiple word

embedding frameworks, such as GloVe and fastText to name a few, have been developed and utilized.

A brief description of the three word embedding frameworks *Word2vec*, *GloVe* and *fastText* is given below. This description touches upon the similarities and differences in embedding creation processes of the three frameworks and gives the reader an insight into the treatment of individual words within these frameworks.

Word2vec

The *Word2vec* framework was proposed in 2013 (Mikolov, Sutskever, et al., 2013), and utilized neural networks to create embeddings and addressed various gaps in the prior word representation methods that incurred high computational costs for large amounts of data. Not only did this method reduce the size of word embedding space, but it densely packed more information within each embedding, pointing at semantic and syntactic regularities (Mikolov, Yih, & Zweig, 2013). This framework, although preserving distances between two words, discarded the statistical properties of the corpus.

Glove

As an approach to address the lack of transparency regarding semantic and syntactic regularities in prior embedding frameworks, the *Global Vectors for Word Representation (GloVe)* framework was conceptualized (Pennington, Socher, & Manning, 2014). This framework combines global matrix factorization with local context window methods (used in *Word2vec*) to learn and create word embeddings. Unlike *Word2vec*, which predicts a given word using neural-networks, *GloVe* uses a frequency-based approach to count the number of times a word has appeared in a context and creates a co-occurrence matrix based on that. In this way, global statistical information of word occurrences within the corpus also play a role when creating word embeddings. According to Pennington et al. (2014), compared to *Word2vec*, *GloVe* showed higher word similarity task accuracy with higher Spearman's rank correlation coefficient between similarity scores and human judgements, as compared to *Word2vec*, even when *GloVe* utilized a corpus less than half the size utilized by *Word2vec*. For the same corpus,

vocabulary, window size and training time *GloVe* claims to achieve faster and more accurate results as compared to *Word2vec*.

fastText

fastText (Bojanowski, Grave, Joulin, & Mikolov, 2017) was proposed in 2016 as an extension of the *Word2vec* model, but with a different treatment of each word during the embedding creation process. Unlike *Word2vec* and *GloVe*, which treat each word as an atomic entity, this framework takes the morphology of words into account and treats each word as a bag of character n -grams. A character n -gram is a connecting sequence of n number of characters in a given word. Each character n -gram has a vector representation and a word composed of multiple character n -grams has a vector representation of the sum of its component character n -grams. This framework claims to create embeddings faster than prior models.

Biases in word embeddings

While word embeddings are very useful in converting raw text input into machine understandable data, they also capture and furthermore amplify the biases present within the input text corpus (Bolukbasi et al., 2016). Due to the context in which certain words are used within human generated data, word embeddings create inappropriate associations between unrelated words, which affects the way a word or an entity defined by that word is treated. Different types of biases have been explored and measured within word embedding frameworks in the preceding literature. These biases can be divided in terms of embedding level biases and corpus level biases.

Embedding level biases

Embedding level biases refer to the biases attached to the word embeddings themselves, affecting fairness in various downstream NLP tasks. These biases associate a particular class of entities with certain inappropriate phenomena more than other classes of the same entities, based on demographic attributes such as gender (binary and non-binary) (Bolukbasi et al., 2016), race (Kiritchenko & Mohammad, 2018), religion (Manzini, Chong, Black, & Tsvetkov, 2019), etc.

Corpus level biases

Prior literature has focused on evaluating bias within word embeddings, but these methods have rarely explored the relationship between the training corpus level biases and their effects on embedding level biases. It has been shown that embedding level biased associations result from the contextual information present within the training corpus (Hu, Zhang, & Zheng, 2016). Corpus level biases include but are not limited to coverage bias, structural bias, lexical bias, etc. (Wagner, Garcia, Jadidi, & Strohmaier, 2015).

Prior research has treated these two types of biases separately, while it should be considered that embedding level biases are generated due to biased training corpora. Thus, it is essential to conduct a study which first focuses on the context, frequency and visibility of words in a given training corpus and then goes on to identify how these features in the training corpus affect the output embedding level biases.

Bias Identification Methods

In order to identify the existence of bias in word embeddings, various quantitative as well as qualitative methods have been developed. Although these methods differ across the type of bias being measured and the type of downstream task being optimized, the inherent structure of conducting bias identification can be divided into five different types of approaches.

- Word Embedding Association Tests
- Analogy based identification
- Bias sub-space identification
- Bias evaluation datasets
- Crowdsourced bias identification

These approaches have been explained in further detail in the next few sections.

Word Embedding Association Tests

Proposed in 2017, this test measures the amount of association between sets of target words with sets of attribute words to determine which target word set is closer to an attribute word set, based on a cosine similarity metric called a WEAT score (Caliskan et al., 2017). This method was developed on the basis of Implicit Association test (Greenwald, McGhee, &

Schwartz, 1998), a measure designed to detect implicit stereotypes in the field of social psychology.

Several extensions of this method have been developed in the form of XWEAT (Lauscher & Glavaš, 2019), which is a multi-lingual and cross-lingual version of WEAT. Another measure, namely Mean Average Cosine Similarity (MAC) was developed in 2019, which extends WEAT testing to a multi-class setting (Manzini et al., 2019). Although WEAT based methods have helped identify specific types of biases in the past, they fail to take into account any implicit biases that aren't listed down in the target or attribute word lists. Moreover, they lead to prejudiced bias measurement techniques, which are heavily impacted by the choice of words in the testing data selected by researcher(s), introducing selectivity in choosing the types of biases they want to explore.

Analogy based identification

Analogy based identification is a qualitative method which uses analogies between words to identify biased or inappropriate associations. An example would be “If a is to b , then c is to ‘?’.” Here “?” will be replaced by the best candidate, chosen according to a metric measuring the similarity between supposed “?” and a candidate word. Several measures have been developed to choose the best candidate in an analogy based task, including CosAdd (Mikolov, Yih, et al., 2013), CosMult (Levy & Goldberg, 2014) and Pairediff (Mikolov, Yih, et al., 2013). These analogies help determine if a certain word is associated with a particular attribute or feature, which might be highly biased or stereotypical. The drawbacks of this method include both selectivity bias and the method's qualitative nature, which leaves much to be desired for when trying to quantify the amount of bias present within a system.

Bias sub-space identification

Another approach to portray the existence of bias in word embeddings involves the creation of a bias sub-space with respect to a particular biased attribute e.g. gender, race, religion, etc. A sub-space B is defined by k orthogonal unit vectors, where if $k=1$, the subspace is simply a direction, as in the case of binary gender, which can be either ‘male’ or ‘female’, defined by a gender direction which is a bias sub-space (Bolukbasi, Chang, Zou, Saligrama, &

Kalai, 2016). In order to demonstrate bias within word embeddings, a gender direction based bias identification approach was developed in 2016. This method measured the distance between supposed ‘gender neutral’ words and ‘gender specific’ words (i.e. binary gender including male and female words). If the distance between a gender-neutral word (e.g. babysitting) was different for two gender specific words with similar meanings and different genders (e.g. grandmother, grandfather), this approach considered it a bias within respective word embeddings (Bolukbasi et al., 2016).

Although this method identifies a bias sub space, it becomes extremely difficult to scale to multi-class settings where biases don’t necessarily occur in binary settings.

Bias evaluation datasets

In the recent years, several researchers have contributed towards bias identification by creating datasets and corpora, which contain specific sentences, which can extract offensive or inappropriate associations between words. One of such datasets is the WinoBias benchmark dataset that was proposed in 2018. This dataset focused on gender bias in coreference resolution, which explored the relationship between certain occupations and the entities corresponding to people, referred by these occupations (Zhao, Wang, Yatskar, Ordonez, & Kai-Wei, 2018). By comparing pro-stereotypical and anti-stereotypical gender-occupation relationships, which were tagged based on occupation statistics, this dataset pointed towards the problem of identifying certain occupations more with one gender than the another.

Crowdsourced bias identification

In order to determine whether certain analogies that are considered biased are actually biased are not, crowdsourced results are also taken into account. Most of the studies in this field utilize crowdsourced data to find whether human perspective aligns with machine generated associations or not. This has been done in several studies including comparing crowdsourced perspectives on several analogy tasks (Bolukbasi et al., 2016) as well as measuring offensive associations (Swinger, De-Arteaga, Thomas Heffernan, Leiserson, & Kalai, 2019), to name a few. Although considered insightful, crowdsourced experiments are highly subjective, depending

upon the type of people taking part in the study. In several studies utilizing this approach, more often than not, no particular control apart from defining superficial measures is employed.

Thus, although various methods have been developed that approach the problem of bias evaluation in different ways, these methods are tainted by the researcher’s selectivity as well as human prejudice. Hence there is a need for robust and meaningful methods that can conduct quantification and qualification of bias in word embeddings without being affected by prior knowledge. These methods should not be based on the effectiveness of generalization of word embedding bias treatment but take specific corpus level properties into account when analyzing bias in word embeddings and how corpus level constraints affect the presence of bias.

Bias Resolution Methods

Once bias is identified within a word embedding, it becomes essential to ensure that bias reduction and even removal is achieved within word embeddings. The models to reduce bias within word embeddings can range from gender direction based modifications (Bolukbasi et al., 2016), word identifier scrubbing from the input text corpus (Prost et al., 2019), swapped data training (Zhao et al., 2019) and a test-time neutralization approach (Zhao et al., 2019). These methods which have been created in prior literature, modify embeddings either during or after the embedding creation process. Described below are the types of methods that have been proposed in the recent literature.

Debiasing as a post-processing task

The concept of debiasing as a post-processing task, i.e. performing bias removal after training word embeddings, was first introduced in 2016 (Bolukbasi et al., 2016). This particular paper specifically targeted gender bias and created a gender direction based post-processing debiasing approach by forcing the trained word vectors for ‘gender-neutral’ terms to lie at zero on the gender axis, removing any supplementary gender-related information within respective embeddings. Here, ‘gender-neutral’ words have been described as words that aren’t associated with a particular gender e.g. ‘money’, ‘determination’, etc. On the other hand, ‘gender-specific’ words have been described as words that are specifically associated with a gender, e.g.

‘grandmother’, ‘queen’, etc. In this study, the researchers handpicked ‘gender-neutral’ and ‘gender-specific’ words.

Although, a novel approach, this study consists of some flaws. To begin with, this method uses a classifier to identify gender-neutral and gender-specific words. Here, classifiers identifying and tagging words makes this method susceptible to the biases present within the classifier itself. In case the classifier misses on tagging a gender-neutral word and tags it as gender-specific, biases won’t be identified within the embeddings. Moreover, this model completely removes gender related information associated with certain words that are useful in medical and social science research (Zhao, Zhou, Li, Wang, & Chang, 2018). Additionally, the debiased results are compared with human perception or social bias over a crowdsourced study, which cannot entirely be considered as a controlled study.

Debiasing as a training task

To address the issues faced by debiasing as a post processing problem, several methods have been developed that introduce debiasing during the training of word embeddings. One such method employs neutralization over stereotypical words by protecting certain attributes and neutralizing other attributes. As long as a list of gender pronouns is provided, this method can target gender bias within word embeddings by protecting the gender attribute, allowing for gender information in several dimensions while removing gender information in other dimensions (Zhao, Zhou, et al., 2018). Although this method has proved to be useful for the task of coreference resolution, the inefficiency of this debiasing method has been portrayed, which states that the words with a strong previous gender bias, even after being debiased through this method, still have the property to be clustered together (Gonen & Goldberg, 2019). This directly leads to a demarcation between sets of words, which have different biases attached to them, still identifiable within the corpus, if plotted. Hence, it is essential that more robust methods be developed that can debias embeddings during the training process.

Corpus augmentation

Corpus augmentation method considers debiasing as a corpus-level task, where word embeddings are trained with a modified corpus to minimize biases. Proposed in 2018, this

method focuses solely on gender debiasing, by taking the original corpus and creating a gender swapped version of the corpus (Zhao et al., 2019). Once the swapping is achieved, the original and the gender swapped corpus variants are combined to train the word embedding framework, in this case ELMo (Peters et al., 2018). This method has been specifically designed for the downstream task of coreference resolution, while using contextualized word embeddings. Moreover, it has shown to be effective decreasing the magnitude of bias amplification from 47.5% to 40.7%, with no major performance losses (Zhao, Wang, Yatskar, Ordonez, & Kai-Wei, 2017).

Although groundbreaking, improvements over these debiasing methods are needed. Several of these debiasing methods have been critiqued, because they provide a specific bias definition, relative to their task and although several debiasing methods ‘succeed’ in debiasing, they merely cover up the bias instead of entirely removing it (Gonen & Goldberg, 2019).

Gaps in Previous Research

Most of the literature on bias identification and reduction in word embeddings is skewed towards identifying and reducing gender biases (based on a binary definition of gender) as compared to handling other types of biases such as racial, religious, age-related biases, with the notable exceptions of (Lauscher & Glavaš, 2019), (Manzini et al., 2019), (Kiritchenko & Mohammad, 2018) and (Rudinger, May, & Van Durme, 2017). Prior methods that evaluate the effectiveness of debiasing are based on different bias evaluation metrics and a wide variety of corpora. This has resulted in the lack of an optimal bias identification and debiasing approach for a given corpus. The definition of bias itself, even if divided into various types, has remained vague. The exact definition of bias is neither consistent nor precise in prior literature.

In order to identify biases, prior research has focused on identifying biased or stereotypical associations between neutral words and attribute words by comparing the similarity metric between lists of neutral words and lists of attribute words (Zhao et al., 2017) (Caliskan et al., 2017). A major flaw in this technique is that these word lists are created by researchers themselves, to pinpoint specific biases or stereotypes present in the trained word embeddings, without considering that other stereotypical word associations might exist within the corpus. This process also introduces results that are heavily influenced by the words picked by the researcher, casting doubt on the effectiveness of the results. Moreover, these word lists used to evaluate

biases aren't modified as one modifies the input training corpora. Thus, variable attributes of the given training corpora are not accounted for, e.g. the size of corpus, word frequencies, context, lexical similarities and differences. This creates an incomparable metric to judge the existence of bias in word embeddings trained from different input corpora and word embedding frameworks.

Moreover, these studies do not take the variability of word usage within a given corpus into account, when defining bias metrics. Given the fluctuating nature of word usage in different contexts, it is important to analyze how bias varies along with these fluctuations. As mentioned by Jawahar & Seddah (2019), "It's important to consider who produced the language, in what context, for what purpose, and make sure that the models are fit to the data."

METHODOLOGY

In this chapter, a brief overview of the methodology employed to conduct identification and evaluation of word associations is described. This work assessed the evolution of word associations over changing corpora and word embedding models, which provided a background for the variability in the nature of biased versus required word associations in word embeddings. This thesis proposed a corpus-adapting word association evaluation approach, in order to account for the semantic and syntactic variations in the corpus. The process began by identifying neutral and attribute words from the raw text corpus. The raw text from the corpus was transformed and cleaned to make it more useful for text modeling and analysis. This processed text was used to train word embeddings, converting words to high dimensional numerical vectors. These vector representations of words were utilized to cluster neutral and attribute words together, using an agglomerative hierarchical clustering approach. Finally, inter-cluster and intra-cluster associations of these clusters were assessed to provide more information about evolution of word associations over changing word embedding models and text corpora. These steps are also listed below:

- Choose text corpus to be analyzed
- Label neutral (proper noun) and attribute (adjective) words from raw text
- Train word embeddings
- Extract word embeddings for neutral (proper noun) and attribute (adjective) words
- Cluster neutral (proper noun) and attribute (adjective) words
- Evaluate variation in word associations between neutral (proper noun) and attribute (adjective) word clusters

The following sections provide a brief overview of the methodology employed to accomplish aforementioned tasks.

Text corpus: Corpus of Historical American English

To conduct an analysis of word embedding associations and biases, the Corpus of Historical American English (Davies, 2015) was utilized. This dataset is the largest structured corpus of historical American English and provides access to the variations in English language

over a time period of 200 years, from 1810s to 2000s (Davies, 2015). The corpus contained more than 400 million words of text from this period, divided across each decade (Davies, 2015). The distribution of raw words and unique words (including proper nouns and adjectives) identified for each decade are noted in table 1.

Table 1 COHA corpus word distribution

Decade	No. of words	Unique words		
		Total Unique words (Proper Nouns + Adjectives)	Proper Nouns	Adjectives
1810s	1,181,022	4,422	2,675	1,747
1820s	6,927,005	10,492	6,695	3,797
1830s	13,773,987	17,167	12,598	4,569
1840s	16,046,854	38,830	29,004	9,826
1850s	16,493,826	19,111	14,134	4,977
1860s	17,125,102	22,294	17,070	5,224
1870s	18,610,160	21,515	16,248	5,267
1880s	20,872,855	23,762	18,359	5,403
1890s	21,183,383	25,783	20,430	5,353
1900s	22,541,232	27,398	21,661	5,737
1910s	22,655,252	27,241	21,334	5,907
1920s	25,632,411	32,057	25,859	6,198
1930s	24,413,247	32,763	26,583	6,180
1940s	24,144,478	33,243	27,139	6,104
1950s	24,398,180	35,037	28,639	6,398
1960s	23,927,982	34,462	28,117	6,345
1970s	23,769,305	35,035	23,245	11,790
1980s	25,178,952	39,639	32,910	6,729
1990s	27,877,340	43,435	36,785	6,650
2000s	29,479,451	47,328	40,905	6,423
TOTAL	406,232,024	571,014	450,390	120,624

In order to analyze how word associations and biases evolved over the period of past 200 years, the text from each decade was used to create word embeddings using five different frameworks: (Word2vec (CBOW), Word2vec (Skip-gram), GloVe, fastText (CBOW), fastText (Skip-gram)).

Labeling neutral and attribute words

In order to limit researcher's bias and to address the variation in text corpora, this thesis aimed to create an automated neutral and attribute word selection approach based on the text

corpus itself. In this approach, proper nouns were used as ‘neutral’ words and adjectives as ‘attribute’ words. Proper nouns were chosen as ‘neutral’ words because they tend to encapsulate multidimensional biases, i.e. bias not limited to a single attribute e.g. gender bias, racial bias, etc. (Kurita, Vyas, Pareek, Black, & Tsvetkov, 2019; Swinger, De-Arteaga, NeilThomasHeffernan, Leiserson, & Kalai, 2018). On the other hand, adjectives were chosen as ‘attribute’ words because when combined with proper nouns, adjectives can be considered attributive (Rind & Tillinghast, 2008).

Within the text corpus, proper nouns and adjectives were identified using part of speech tagging through StanfordNLP pipeline. This is described in more detail in the following sections.

Part of Speech Tagging

Part of speech (POS) tagging is the process of tagging ‘parts-of-speech’ in the given text corpus. The software read the given text and assigned a part of speech to each word in the given text. In this thesis, the part-of-speech tagging process was implemented using the StanfordNLP Pipeline (v0.2.0). This pipeline utilized a highway BiLSTM, concatenating inputs from pretrained word embeddings, trainable word embeddings and a character level embedding generated from a unidirectional LSTM over word characters (Qi, Dozat, Zhang, & Manning, 2018). Once tagging was completed, ‘JJ’ and ‘PROPN’ tagged words were filtered, where ‘JJ’ was the tag given to adjectives, and ‘PROPN’ was the tag given to proper nouns, in order to select adjectives and proper nouns from the given text.

Tagger Validity

It was noted that the aforementioned tagger being used to tag proper nouns and adjectives did not portray a 100% tagging accuracy. Thus, in order to check for the extent of how accurately proper nouns and adjectives were being tagged, 20 tagged proper nouns and adjectives each for text data from each of the twenty decades were randomly selected and checked if they had been correctly tagged. Overall, the average accuracy of the tagger in tagging proper nouns was 69.4% and in tagging adjectives was 63.2%. Thus, overall, proper nouns were tagged more accurately as compared adjectives. It was observed that words with roots from different languages, misspelled words, words incorrectly combined together using hyphens and special

characters and acronyms were being incorrectly tagged as proper nouns and adjectives in multiple cases.

Training word embeddings

The raw text served as an input to the given word embedding framework. In this process, each word was transformed into a multidimensional vector and projected in a multidimensional vector-space. Words with similar meanings were expected to have similar representations and more probable to lie closer to each other in the vector-space as compared to words with different meanings.

When using the *Word2vec* framework, each word was taken as an individual token and a vector representation was created. When using the *GloVe* framework, not only were the local word contexts but global statistical information of words was also taken into account to create vector representations. When using the *fasttext* framework, each word was divided into a bag of character n-grams, where each character n-gram was assigned a vector representation. Thus, the vector representation of a word comprising of multiple character n-grams was the sum of the vector representations of its component character n-grams.

Table 2 Model Hyperparameters

Parameter	window	model	alpha	epochs
Word2vec (CBOW)	5	sg = 0 [cbow]	0.025	1
Word2vec (Skip-gram)	5	sg = 1 [skip-gram]	0.025	1
GloVe	5	-	0.025	1
fastText (CBOW)	5	“cbow”	0.025	1
fastText (Skip-gram)	5	“skipgram”	0.025	1

In this thesis, two different approaches were used to train *Word2vec* and *fastText* embeddings, i.e. continuous bag of words (CBOW) and Skip-gram. For the CBOW approach, the network tries to predict which word is most likely given its context. Words that are equally likely to appear can be interpreted as having a shared dimension. For the Skip-gram approach, the network tries to predict neighboring words (context) which are most likely, given the current word.

In order to keep the comparison between different word embedding models consistent, the hyperparameters were kept uniform, as shown in table 2. Thus, after creating the word embeddings using these five different frameworks, the vector representations of each word within the corpus was obtained.

Creating neutral and attribute word lists/clusters

In order to create lists/clusters of neutral words (proper nouns) and attribute words (adjectives), a clustering approach was utilized. Clustering is the process of grouping together similar entities, such that entities in the same group are more similar than the entities in different groups. Here, the entities are words, which were clustered based on their vector representations, i.e. words with similar vector representations were clustered together, as compared to words with different vector representations. Clustering has been utilized to create these word groups because it is highly unsupervised and minimized any explicit researcher's bias when creating proper noun and adjective word lists.

The clustering approach used for this task is agglomerative hierarchical clustering. This approach is selected over other clustering approaches because not only did it produce a hierarchy partition of the given text data, providing information at each step of sub-cluster creation, but it also provided a deterministic and more flexible clustering mechanism (Aggarwal, 2013). Moreover, it has been extensively used in multiple information retrieval and text mining tasks (Wei, 2019). Agglomerative clustering adopts a 'bottom-up' approach, starting with each entity represented as an individual cluster, which are merged together after each iteration depending on various distance measures (Wei, 2019). Agglomerative clustering is further divided into different types based on the cluster distance measure used to combine entities to form cluster. These measures are single link, complete link, average link and ward's link. Single link clustering measures the shortest distance between two points in each cluster, complete link clustering measures the longest distance between two points in each cluster, average link clustering measures the average distance between each point in each cluster and ward linkage measures distance between clusters by minimizing sum of squared differences within all clusters (Aggarwal, 2013).

This thesis utilized ward linkage based agglomerative hierarchical clustering. This clustering method accounted for the merging cost of combining a pair of clusters, uncovering

clusters that might not necessarily be round or similar in diameter. The measure of distance used to compute the proximity was $1 - \text{cosine similarity}$. This *cosine similarity* measured the cosine of the angle formed between vector representations of two entities in the vector-space, given by the word embedding frameworks discussed above, bound by the range [0,1]. Thus, the more similar two words were, the smaller the angle between their vector representations and the closer the cosine of that angle, i.e. *cosine similarity*, was to 1.

In order to account for the lexical richness and variability of the input text corpora and to create a baseline for the number of clusters created for each corpus, across every word embedding model, the input proper nouns and adjectives were classified into eight clusters. Motivated by the Plutchik's wheel of emotions (Plutchik, 1980), presumably, these eight clusters would represent emotion themes and shed light on distinctions between certain proper nouns and adjectives associated with them. The wheel of emotions specifies eight primary bipolar emotions: joy versus sadness; anger versus fear; trust versus disgust; and surprise versus anticipation.

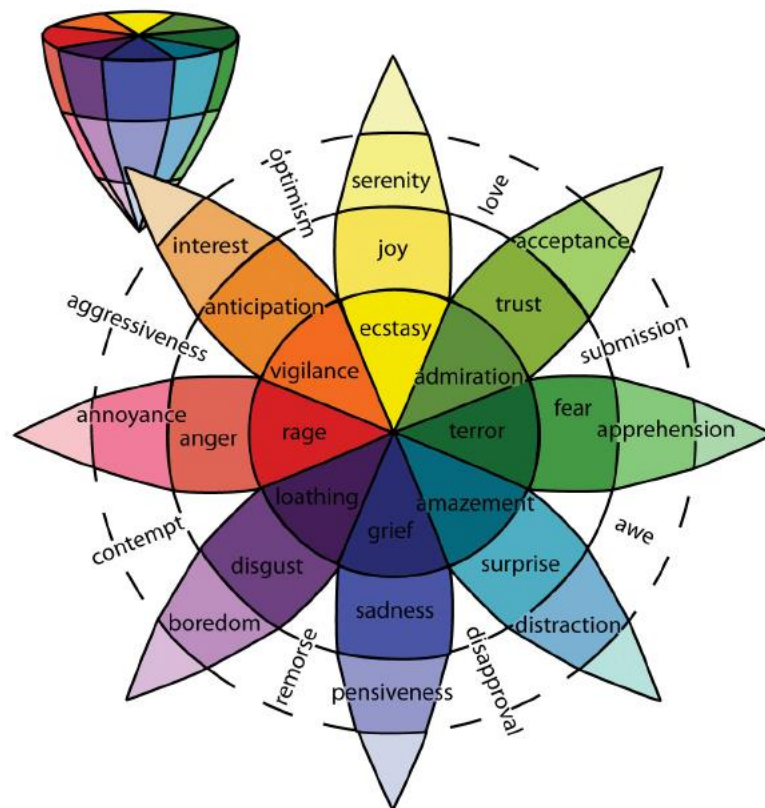


Figure 1 Plutchik's wheel of emotions (Maupomé & Isyutina, 2013)

Measuring word associations

Once neutral words (proper nouns) and attribute words (adjectives) were tagged, embedded and clustered to form representative groups, neutral-attribute (proper noun-adjective) word associations were measured by assessing the quality of created clusters using quantitative and qualitative approaches.

Quantitative Measures

Once the clusters were created, to measure word associations and identify how closely knit or spread apart the proper noun-adjective word representations were within the created clusters, word associations were evaluated using inter-cluster and intra-cluster analyses. While inter-cluster analysis evaluated the quality of clustering by assessing the overall distribution of clusters, intra-cluster analysis evaluated the quality of clustering by assessing the distribution of elements (words) within created clusters.

Inter-cluster analysis

In order to analyze whether words from a given text corpus were creating well separated clusters (distinctive word associations) or loosely defined clusters (vague word associations), a cluster validity metric i.e. *Dunn's Index* was computed. This metric was utilized to assess cluster validity and aims to identify tightly knit and well-separated clusters (Xu, 2015). The equation to calculate Dunn's Index is given below:

$$Dunn's\ Index = \frac{Minimum\ inter\ cluster\ distance}{Maximum\ cluster\ diameter}$$

Thus, for K given clusters, *Dunn's Index* will be calculated as $Du(K)$ as follows;

$$Du(K) = \min_{i=1,\dots,K} \left(\min_{j=i+1,\dots,K} \left(\frac{D(C_i, C_j)}{\max_{l=1,\dots,K} diam(C_l)} \right) \right)$$

where

$$D(C_i, C_j) = \min_{x \in C_i, y \in C_j} D(x, y)$$

$$diam(C_i) = \max_{xy \in C_i} D(x, y)$$

Hence, a large value of Dunn's Index indicated that the inter-cluster distance was maximized, and cluster diameter was minimized, to form compact and well-separated clusters. On the other hand, a low value of Dunn's Index indicated loosely defined and less crisp clusters.

Intra-cluster analysis

To assess the distribution of words within each cluster and to evaluate the similarities and differences in word associations captured by different word embeddings, two different types of cluster-level data were captured.

Distribution of words across clusters

By plotting the distribution of words across clusters for each word embedding model it was revealed whether words were uniformly divided into clusters, signifying distinctive word associations, or whether these words were divided primarily into one or two clusters, showing uneven distribution of words across clusters, signifying lack of distinction in word associations.

Jaccard Similarity between different word embedding model clusters

In order to find if there was any similarity in the way clustering was being done for the five given word embedding models, it was crucial to compare the words being grouped together for each word embedding model. Thus, the clusters created for every word embedding model were compared with the clusters created for the other four word embedding models. To compute this, average Jaccard Similarity between the clusters for each pair of word embedding models was calculated. Average Jaccard Similarity between cluster A and cluster B was calculated as follows.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

In this thesis, to compare clustering done using two word embedding models W_A and W_B , the Jaccard Similarity metric was extended over to the respective sets of clusters, $C_A : [C_{A1}, C_{A2}, \dots, C_{A8}]$ and $C_B : [C_{B1}, C_{B2}, \dots, C_{B8}]$.

$$J(C_A, C_B) = \frac{J(C_{A1}, C_{B1}) + J(C_{A1}, C_{B2}) + \dots + J(C_{A8}, C_{B8})}{64}$$

Finally, the average Jaccard Similarity between clusters created by word embedding models W_A and W_B , over 20 decades (1810s – 2000s) was calculated.

$$Average J(C_A, C_B)_{1810s-2000s} = \frac{J(C_A, C_B)_{1810s} + J(C_A, C_B)_{1820s} + \dots + J(C_A, C_B)_{2000s}}{20}$$

Similarly, measuring the Jaccard Similarity between clusters of difference word embedding models evaluated the similarities and differences in word associations being captured by each word embedding model.

Qualitative Measures

Apart from reporting the quantitative measures of similarities and differences between word associations captured by word embedding models, this work also evaluated the variations in semantic and syntactic regularities being captured by every word embedding model, over a period of twenty decades.

Semantic regularities were evaluated by observing the similarity in meanings of the words being clustered together. Syntactic regularities on the other hand, were evaluated by observing the morphological similarities of the string formats of the words being clustered together.

Additionally, this work also explored how incorrectly tagged proper nouns and adjectives were encoded as word embeddings and distinguished from correctly tagged proper nouns and adjectives respectively.

Evaluation Framework

The systematic evaluation of deviations in word associations was done using the evaluation framework described in table 3.

Table 3 Evaluation framework

Evaluation	Fixed	Variable	Observations
Evaluation I	Text corpus	Word embedding model	- Dunn's Index (DI) - Distribution of word within clusters - Jaccard Similarity - Semantic and Syntactic regularities
Evaluation II	Word embedding model	Text corpus	- Dunn's Index (DI) - Corpus Size

Evaluation I: Variation in word associations over varying word embedding models

In this analysis, the text corpus was fixed and used to train five different word embedding frameworks, Word2vec (CBOW), Word2vec (Skip-gram), GloVe, fastText (CBOW) and fastText (Skip-gram). The variation in word associations with respect to changing word embedding models was computed by measuring changes in Dunn's Index, distribution of words within clusters, Jaccard similarity between clusters of different word embeddings and changes in semantic and syntactic regularities captured by every word embedding model.

Table 4 Dunn's Index over varying word embedding models and fixed text corpus

Dunn's Index					
Decade	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fasttext (Skip-gram)
Text corpus					

Evaluation II: Variation in word associations over varying text corpora

In this analysis, the word embedding model was fixed and trained over varying text corpora. Then, the differences in word associations were compared by measuring changes in Dunn's Index with respect to quantitative parameters of the corpus i.e. corpus size changed, over each decade.

Table 5 Dunn's Index over varying text corpora and fixed word embedding model

Decade	Dunn's Index		
	Word Embedding model 1	...	... Word Embedding model 5
1810s			
...			
2000s			

Thus, by calculating the aforementioned measures over varying text corpora and word embedding frameworks, this thesis examined how word associations develop across different

word embeddings, and whether there exist any key properties of a given word embedding model that promote certain word associations more than others, generating bias.

RESULTS

The results from clustering neutral words (proper nouns) and attribute words (adjectives) include the analysis of 800 clusters, across 20 decades of textual data, divided over 5 different types of word embeddings. These results are divided into three sections. The first section reports the quantitative variation in Dunn’s Index over different word embedding models and training text corpora. The second section reports the similarities and differences in clustering patterns, done for different word embedding models. The third section reports a qualitative analysis of the elements (proper nouns and adjectives) being clustered together, to reveal whether semantically or syntactically similar clusters are being formed.

Evaluation I: Variation in word associations over varying word embedding models

Change in Dunn’s Index

Variations in Dunn’s Index (DI) of clusters made using different word embedding models, over a period of twenty decades, is shown in figure 2. Average Dunn’s Index and standard deviation over twenty decades is also reported in table 6.

Table 6 Dunn's Index for word embedding model clusters, N = 20

Model	<i>M</i> (DI)	<i>SD</i> (DI)
Word2vec (CBOW)	0.021	0.006
Word2vec (Skip-gram)	0.063	0.011
GloVe	0.012	0.006
fastText (CBOW)	0.004	0.002
fasttext (Skip-gram)	0.023	0.006

Note: M and SD are used to represent mean and standard deviation respectively

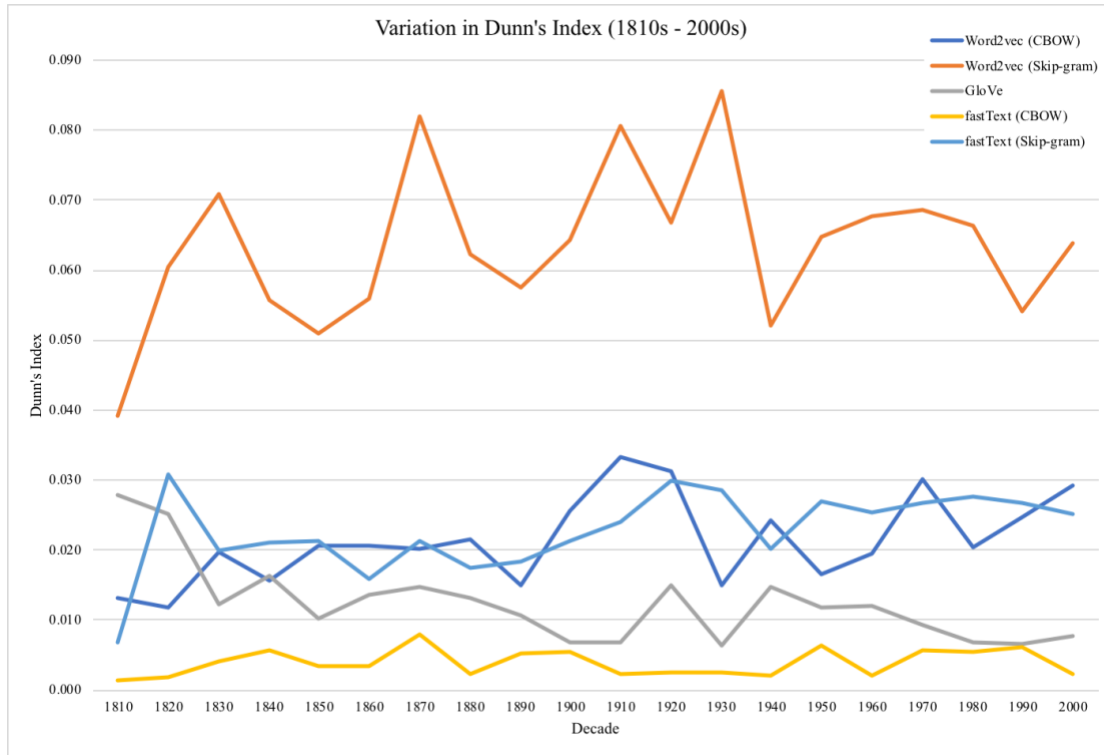


Figure 2 Plot of Dunn's Index over 20 decades (1810s - 2000s)

A higher Dunn's Index represents compact and well-separated clusters, while a lower Dunn's Index represents loosely defined and less separated clusters.

It was observed that Word2vec (Skip-gram) embeddings consistently created the most compact and well separated clusters, across all twenty decades of data. Thus, words embedded using Word2vec (Skip-gram) model formed word associations that were more distinguishable than the word associations formed for words embedded using the other four word embedding models. The variation of Dunn's Index across different text corpora was higher for Word2vec (Skip-gram) embeddings as compared to other models, showing that the quality of clustering changed a lot more with changes in the training corpora, for Word2vec (Skip-gram) embeddings. Comparatively, Word2vec (CBOW) embeddings consistently portrayed a lower quality of clusters, showing that the words embedded using Word2vec (CBOW) model formed word associations that were less distinguishable than those formed for words embedded using Word2vec (Skip-gram) model.

This distinction between clustering done using CBOW and Skip-gram embedding architectures was also visible for fastText word embedding models. On an average as well as

across all the twenty decades consistently, fastText (Skip-gram) models captured more distinguishable word associations as compared to fastText (CBOW) models. Hence, the Skip-gram architecture was more sensitive towards capturing the similarity and dissimilarity between words, as compared to the CBOW architecture.

Moreover, when comparing standard deviations of Dunn's Indices across models in table 6, it was seen that Dunn's index for Skip-gram models deviated more with changes in text corpora, as compared to the CBOW models. Thus, as the text corpus changed, Skip-gram models showed more variation in the quality of word associations being captured, i.e. the word associations being captured by the Skip-gram model were more affected by the training corpora as compared to the CBOW architecture.

Compared to the other four word embedding models, GloVe word embedding clusters portrayed a degrading performance in terms of capturing word associations, as in our text corpora changed across decades, by showing a consistently low value of the Dunn's Index.

Overall, the word embedding model that captured the least information regarding word associations turned out to be fastText (CBOW), which created clusters of the lowest quality evaluated in terms of Dunn's index.

These observations revealed that Word2vec embeddings, especially Word2vec (Skip-gram), captured the most information regarding similarities and dissimilarities between words as compared to GloVe and fastText. When considering the different architectures, CBOW and Skip-gram, it can be seen that word embedding models utilizing the Skip-gram architecture created more distinguishable clusters, revealing that they captured more information about the distinctions between different words, as compared to CBOW. Moreover, it can be seen that Skip-gram architecture was affected more by changes in the training text corpora, as compared to CBOW, which was reflected by the higher standard deviations for Dunn's indices, across text corpora, for these models. This showed that the sensitivity of Skip-gram models towards lexical regularities was higher than CBOW models.

Distribution of words within clusters

The distribution of words across each cluster showed how the clustering patterns changed with respect to the word embedding models, while controlling for the training text corpora. The distribution across all five word embedding models, is shown in Fig. 4.2. It should be noted that

if words were evenly distributed within clusters, each cluster would contain a similar proportion of words, leading to a narrower spread of words across clusters. Consequently, if words were unevenly distributed within clusters, some clusters would contain a much higher proportion of words than other clusters, leading to a wider spread of words across clusters.

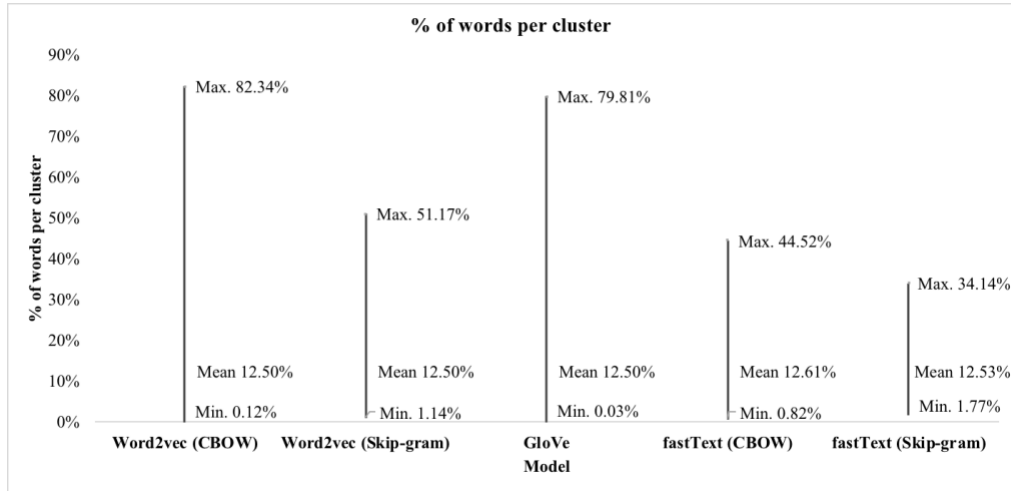


Figure 3 Distribution of words across embedding clusters

It was observed that Word2vec (Skip-gram) and fastText (CBOW & Skip-gram) models created more uniform clusters as compared to Word2vec (CBOW) and GloVe models. Overall, the most uniform clustering was achieved by fastText (Skip-gram) models, followed by fastText (CBOW) and Word2vec (Skip-gram).

For both Word2vec (CBOW) and GloVe models, some clusters contained more than 75% of the words being clustered, essentially grouping most of the words together in a limited number of clusters, creating highly non uniform clusters. This showed a lack of distinguishable word associations for GloVe and Word2vec (CBOW) word embedding models, as compared to other word embedding models.

When comparing these five words embedding models, fastText word embeddings tended to create the most uniform clusters, as compared to Word2vec and GloVe embeddings, over the same training corpora. This was followed by Word2vec embeddings, which comparatively created more uniform clusters as compared to GloVe embeddings. When comparing the CBOW and Skip-gram architectures, it was revealed that for both Word2vec and fastText models, the Skip-gram architecture led to more uniform clustering as compared to the CBOW architecture.

Thus, even though a given text corpus might contain a common set of themes, these themes were captured differently across the five words embedding models. This also revealed that the word associations encoded by the word embeddings differed across the different word embedding models, which captured different lexical regularities from the same corpus.

Jaccard Similarity between clusters from difference word embeddings

The average Jaccard Similarity between clusters created for each pair of word embedding models, over twenty decades, is shown in table 7. Jaccard Similarity between word embedding clusters for each decade have been included in Appendix B.

Table 7 Average Jaccard Similarity between word embedding models (1810s – 2000s)

Model	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	-	-	-	-
Word2vec (Skip-gram)	0.38	1.00	-	-	-
GloVe	0.44	0.36	1.00	-	-
fastText (CBOW)	0.34	0.46	0.35	1.00	-
fastText (Skip-gram)	0.39	0.50	0.41	0.62	1.00

These results reveal that the Word2vec (CBOW) embeddings captured more similar word associations with GloVe embeddings than Word2vec (Skip-gram) embeddings. Hence, even though both the Word2vec models have a common architecture, the difference in considering the word context to create the embeddings had a direct impact on the word associations being captured.

It can also be seen that Word2vec (Skip-gram) and fastText (Skip-gram & CBOW) embeddings created very similar clusters as compared to Word2vec (Skip-gram) and Word2vec (CBOW) embeddings. Additionally, fastText (CBOW) and fastText (Skip-gram) models created the most similar clusters, across pairs of all five word embedding models.

The distinction between CBOW and Skip-gram architecture was more visible in the clustering patterns of Word2vec as compared to fastText. This signified that word associations captured by Word2vec models were more affected by the network architecture used to derive the

embedding, i.e. whether the word had been predicted given a context, versus the context was predicted given a word, as compared to fastText models.

Qualitative analysis: Semantic and syntactic regularities across word embeddings

In this section, the clusters being formed have been qualitatively analyzed. Prior to this analysis, for every decade, five different word embedding models were trained and used to map proper nouns and adjectives to a vector space. These proper nouns and adjectives were clustered together based on word associations learnt by the respective word embedding model. For each of the five word embedding models, these proper nouns and adjectives were divided into eight distinct clusters, i.e. 40 total clusters are formed. For every decade and word embedding model, clusters were represented by the words closest to the cluster center, aiding in the qualitative analysis of variations in the words being clustered, across different word embedding models.

Table 8 Rare words clustered together (Word2vec (CBOW), 1910s, Cluster 3)

Cluster	Word	Frequency
['pettilow', 'dijon', 'melrose', 'camembert', 'ashgrove', 'lenman', 'wrenfield', 'churchman', 'papist', 'unprofessional', 'dry-mouthed', 'dimmest', 'stoutish', 'undifferentiated', 'unwell', 'humoured', 'extra-heavy'] *(Proper noun, Adjective)	pettilow	2
	dijon	4
	melrose	9
	camembert	2
	ashgrove	7
	lenman	10
	wrenfield	10
	churchman	11
	papist	5
	unprofessional	15
	dry-mouthed	1
	dimmest	7
	stoutish	6
	undifferentiated	10
	unwell	9
	humoured	5
	extra-heavy	1

For Word2vec (CBOW and Skip-gram) embeddings, it was observed for multiple instances that unique and less frequent words were being clustered together, as shown in table 8 and table 9. Thus, less frequent proper nouns were being clustered with less frequent adjectives. This shows that the frequency of the occurrence of a word within a corpus affected its embedding representation in a way, that led to stronger word associations between words which might not be semantically similar but shared the property of rarity within a corpus.

Table 9 Rare words clustered together (Word2vec (Skip-gram), 1940s, Cluster 3)

Cluster	Word	Frequency
['stechers', 'gunnar', 'ambergris', 'perce', 'wynnewood', 'grantland', 'brittain', 'fagan', 'stasia', 'nothinghappened', 'heraldic', 'disobedient', 'natured', 'constipated', 'underhand', 'cheeked', 'larval', 'withhis', 'saleable'] *(Proper noun, Adjective)	stechers	3
	gunnar	7
	ambergris	6
	perce	5
	wynnewood	13
	grantland	4
	brittain	4
	fagan	14
	stasia	8
	nothinghappened	4
	heraldic	15
	disobedient	10
	natured	5
	constipated	14
	underhand	6
	cheeked	5
	larval	5
	withhis	3
	saleable	4

Moreover, names of people, places and groups were more distinguishable in Word2vec word embeddings, followed by GloVe and fastText embeddings. Some examples of names of people, places and groups, clustered together are in table 10, table 11 and table 12.

Table 10 Person names clustered together (Word2vec embeddings)

1930s	1930s	1960s	1970s
['yancey', 'gee', 'granny', 'jack', 'robinson', 'ed', 'tom', 'scott', 'sam']	['robinson', 'winfield', 'scott', 'david', 'payne', 'elias', 'lincoln', 'george', 'jefferson']	['tom', 'charlie', 'jesus', 'ben', 'jessup', 'leroy', 'bible', 'perkins', 'sylveste']	['otto', 'sophie', 'god', 'charlie', 'ruth', 'garland', 'judy', 'myron', 'flo']

Table 11 Place/Location names clustered together (Word2vec embeddings)

1930s	1960s	1970s	1980s
['kansas', 'camp', 'territory', 'oklahoma', 'georgia', 'columbus', 'washington', 'wagon', 'california']	['mount', 'colorado', 'park', 'sunday', 'wednesday', 'los', 'angeles', 'company', 'christmas']	['brooklyn', 'heights', 'street', 'broadway', 'flynders', 'city', 'bush', 'avenue', 'island']	['york', 'landing', 'friday', 'korea', 'road', 'holiday', 'orchestra', 'washington', 'delaware']

Table 12 Group names clustered together (Word2vec embeddings)

1960s	1980s	1990s	1910s
['elders', 'indians', 'insurance', 'americans', 'services', 'allies', 'germans', 'years', 'ages']	['states', 'adults', 'bond', 'may', 'americans', 'europe', 'christians', 'will', 'chiefs']	['dragons', 'workers', 'dogs', 'guards', 'brothers', 'words', 'writers', 'sisters', 'fathers']	['italians', 'ages', 'russians', 'asia', 'girls', 'people', 'books', 'virtues', 'greeks']

Table 13 Syntactically similar words clustered (fastText)

Tagged names	['ed', 'embalmed', 'den', 'tailed', 'weed', 'fell', 'heard', 'ted', 'staked']	['building', 'flying', 'king', 'ring', 'riding', 'dang', 'bombing', 'pal', 'bunting']	['den', 'teen', 'ned', 'en', 'grant', 'ed', 'ted', 'gran', 'reed']
Tagged adjectives	['scared', 'amused', 'embarrassed', 'relaxed', 'mottled', 'excited', 'subdued', 'surprised', 'thrilled', 'worried']	['amusing', 'enchanted', 'troubling', 'upsetting', 'distressing', 'exhausting', 'upcoming', 'fascinating', 'disgusting', 'exciting']	['excited', 'worried', 'outstretched', 'hotheaded', 'pleased', 'frightened', 'fatigued', 'astounded', 'annoyed', 'impressed']

Another pattern that emerged over repeated embedding training over different corpora was that fastText embeddings (CBOW and Skip-gram) captured syntactic similarities far more often than other word embedding models, as shown in table 13. This result may be expected due to the architecture of fastText.

Adjectives used in the similar context were clustered very well in Word2vec (Skip-gram) word embeddings. Adjectives describing physical characteristics of a person, human propensity, amount, etc., were distinguishable in the form of different clusters in Word2vec (Skip-gram) word embeddings, followed by Word2vec (CBOW) embeddings.

Moreover, prepositions and verbs misidentified as adjectives by the StanfordNLP Pipeline tended to group together when embedded using Word2vec (Skip-gram) and Word2vec (CBOW) models. This shows that Word2vec embeddings captured the difference between the usage of adjectives, prepositions, verbs, etc., more than GloVe and fastText word embedding models. It should also be noted that words embedded using GloVe model also captured distinction between misidentified proper nouns and actual proper nouns.

The frequency of incorrectly tagged proper nouns increased as the size of the corpus being tagged increased. This was an artifact of the StanfordNLP Pipeline POS tagger. These incorrectly tagged words are also reflected in the cluster representations, as shown in table 12, where words like ‘adults’, ‘states’, ‘years’, etc. were tagged by the tagger as proper nouns, while these are examples of common nouns.

While several instances of clustering showcase clusters containing words representing similar emotion themes, as described in the Plutchik’s wheel of emotions, it is also seen that adjectives get clustered based on their usage within language and the context in which they are used, e.g. describing human propensity, describing physical characteristics, describing a group, place, etc. Moreover, certain instances observe words describing emotions on opposite ends of the Plutchik’s wheel of emotions, e.g. ‘Loathing’ and ‘Admiration’, being clustered together. Thus, semantic clusters are formed in such cases, not necessarily replicating Plutchik’s wheel of emotions.

Evaluation II: Variation in word associations over varying text corpora

Change in Dunn's Index

Over different decades, the quality of clusters varied for the given five models, as shown in table 14. The plots shown in figure 4 also visualize the differences in clustering quality across different word embedding models, for the same training corpora.

Table 14 Variation in Dunn's Index over decades

Decade	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fasttext (Skip-gram)
1810s	0.021	0.052	0.047	0.002	0.006
1820s	0.032	0.072	0.025	0.008	0.039
1830s	0.027	0.084	0.014	0.008	0.061
1840s	0.041	0.061	0.015	0.010	0.068
1850s	0.032	0.051	0.010	0.010	0.068
1860s	0.044	0.062	0.015	0.010	0.061
1870s	0.034	0.071	0.016	0.011	0.079
1880s	0.038	0.070	0.012	0.011	0.065
1890s	0.031	0.067	0.013	0.013	0.070
1900s	0.036	0.076	0.007	0.013	0.067
1910s	0.029	0.075	0.007	0.013	0.065
1920s	0.026	0.069	0.013	0.013	0.073
1910s	0.029	0.075	0.007	0.013	0.065
1930s	0.043	0.071	0.007	0.014	0.067
1940s	0.027	0.089	0.012	0.013	0.080
1950s	0.038	0.091	0.011	0.013	0.067
1960s	0.034	0.067	0.014	0.013	0.068
1970s	0.026	0.086	0.009	0.012	0.070
1980s	0.034	0.069	0.007	0.013	0.070
1990s	0.027	0.078	0.008	0.014	0.068
2000s	0.025	0.073	0.008	0.015	0.062

It was observed that the quality of word clusters varied across different word embedding models, even when trained on the same training corpus. Across all five word embedding models, it can be seen that Word2vec (Skip-gram) model architecture consistently outperformed the other

four models, in terms of capturing distinctive word associations from a given text. Moreover, the clustering quality improved for all of the models as the size of the corpus increased, apart from GloVe, whose clustering quality degraded as the text corpus became larger.

Overall, it can be observed that the Skip-gram architecture created better quality clusters, i.e. Skip-gram word embeddings were able to capture more distinctive word associations, capturing a higher degree of similarity and dissimilarity between words within the vector space, as compared to the CBOW word embeddings.

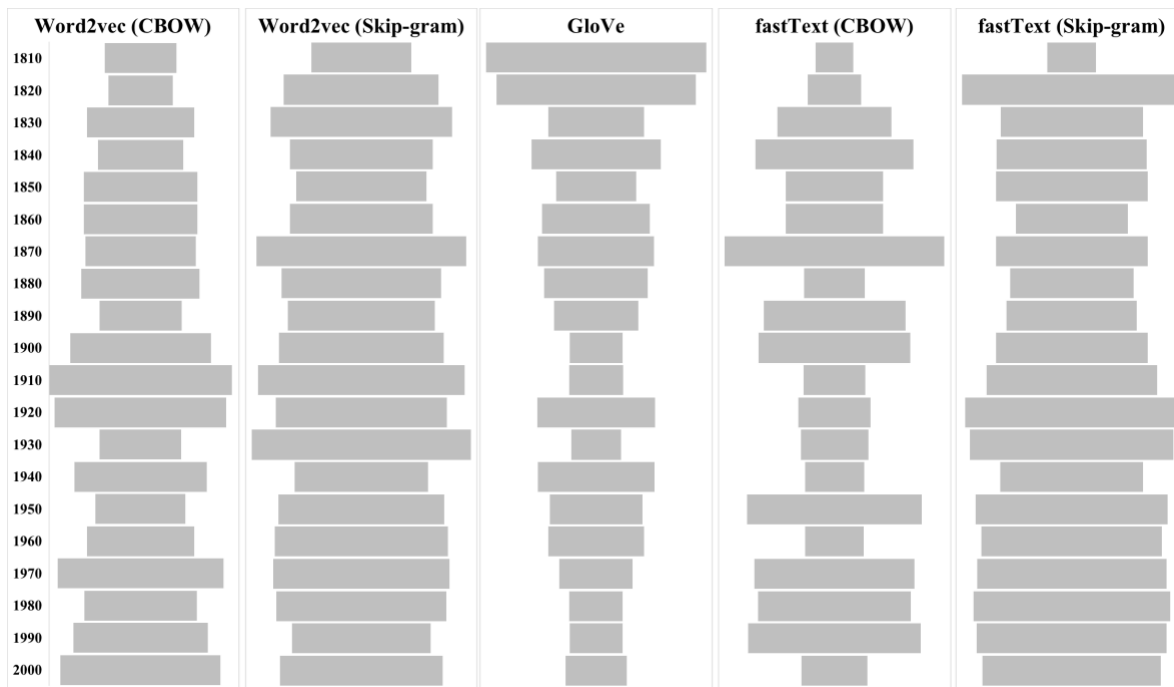


Figure 4 Variation in Dunn's Index across word embeddings (1810s -2000s)

Future Work

The work in this thesis is a first step towards comparing various embedding models across corpora. Future work can include several additions to this framework, where, instead of using an automated part-of-speech tagger to tag adjectives and proper nouns, a corpus specific taxonomy of adjectives can first be created, which will assist in determining the baseline for the number of word clusters that need to be created. This will also improve the accuracy and specificity of the analysis and help reveal more significant and meaningful word associations between neutral words (proper nouns) and attribute words (adjectives). It is also crucial to account for the vast variations in the vocabulary from 1810s to 2000s, which ranges from rare

archaic English words to recently created slang words. This will also improve the accuracy of the words being tagged and the embeddings being generated by capturing the context each rare word is actually used in.

Moreover, more research needs to be conducted on the qualitative analysis of words being clustered together. This can be achieved by subject matter experts annotating each cluster, which can help create a rubric to identify all different types of biases a particular word association might represent.

CONCLUSION

This work created an automated framework for the analysis of word associations in diachronic word embeddings, for five different word embedding frameworks, over 20 decades of textual data. For each text corpus, part-of-speech tags were created using StanfordNLP Pipeline POS tagger and five different word embedding models were trained (Word2vec (CBOW), Word2vec (Skip-gram), GloVe, fastText (CBOW), fastText (Skip-gram)). The words tagged as proper nouns and adjectives were extracted, along with their respective word embedding representations. These proper noun and adjective word embeddings were clustered using ward linkage based agglomerative hierarchical clustering. Finally, the resulting clusters were qualitatively and quantitatively analyzed, with respect to changes in word embedding models and training corpora, to reveal any significant changes in word associations.

For the given corpora, it was observed that on an average, Word2vec (Skip-gram) embeddings captured the most distinctive word associations, followed by fastText (Skip-gram), Word2vec (CBOW), GloVe and fastText (CBOW). Moreover, word associations captured by Word2vec (CBOW) embeddings are most similar to GloVe embeddings. This is quantitatively corroborated by the Jaccard similarity metric calculated between these word embedding models. This might be a result of uneven distribution of elements within clusters, as shown in the plot of distribution of words across clusters, showing that Word2vec (CBOW) and GloVe embeddings show weaker and less distinctive word associations as compared to Word2vec (Skip-gram) embeddings. Clusters created by Word2vec (Skip-gram) and fastText (Skip-gram) embeddings also portray a high similarity, showing that similar word associations are being captured by their respective word embeddings. This might be attributed to the similarity in their approach of using one word to predict the context, while creating word embeddings. Additionally, fastText (CBOW) and fastText (Skip-gram) embeddings show a high similarity in word associations, sharing the same embedding architecture but different network approaches. Word associations between neutral and attribute words embedded using Word2vec (Skip-gram) and fastText (CBOW) models significantly strengthened as the size of the corpus increased. On the other hand, word associations between neutral and attribute words embedded using GloVe model significantly deteriorated as the size of the corpus increased. No significant relationship was

observed for word associations between neutral and attribute words embedded by other models (Word2vec (CBOW), fastText (Skip-gram)) as the size of the corpus increased.

The methods of creating a rubric which can distinguish between a biased word association and a required word association are still in infancy and need improvement. This work presents an initial framework to objectively analyze word associations and also presents a comparison of the word associations generated across different word embedding models, trained under controlled hyperparameters and input training text.

REFERENCES

- Aggarwal, C. (2013). *Data Clustering* (1st edition ed.): Chapman and Hall/CRC.
- Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, 3(Feb), 1137-1155.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135-146.
- Bolukbasi, T., Kai-Wei, C., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *arXiv.org*.
- Caliskan, A., Bryson, J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *arXiv.org*, 356(6334). doi:10.1126/science.aal4230
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2), 153-163.
- Collobert, R., & Weston, J. (2008). *A unified architecture for natural language processing: Deep neural networks with multitask learning*. Paper presented at the Proceedings of the 25th international conference on Machine learning.
- Davies, M. (2015). *Corpus of Historical American English (COHA)* [linguistic corpora]. Retrieved from: <https://doi.org/10.7910/DVN/8SRSYK>
- Gonen, H., & Goldberg, Y. (2019). Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But do not Remove Them.
- Gordon, J., & Van Durme, B. (2013). *Reporting bias and knowledge acquisition*. Paper presented at the 2013 Workshop on Automated Knowledge Base Construction, AKBC 2013 - Co-located with CIKM 2013, October 27, 2013 - October 28, 2013, San Francisco, CA, United states.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6), 1464.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10(2-3), 146-162.
- Hu, W., Zhang, J., & Zheng, N. (2016). *Different contexts lead to different word embeddings*. Paper presented at the 26th International Conference on Computational Linguistics, COLING 2016, December 11, 2016 - December 16, 2016, Osaka, Japan.
- Kiritchenko, S., & Mohammad, S. (2018). Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. *arXiv.org*.
- Lauscher, A., & Glavaš, G. (2019). *Are We Consistently Biased? Multidimensional Analysis of Biases in Distributional Word Vectors*. Paper presented at the Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (* SEM 2019).
- Levy, O., & Goldberg, Y. (2014). *Linguistic regularities in sparse and explicit word representations*. Paper presented at the Proceedings of the eighteenth conference on computational natural language learning.

- Manzini, T., Chong, L. Y., Black, A. W., & Tsvetkov, Y. (2019). *Black is to Criminal as Caucasian is to Police: Detecting and Removing Multiclass Bias in Word Embeddings*. Paper presented at the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers).
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality*. Paper presented at the Advances in neural information processing systems.
- Mikolov, T., Yih, W.-t., & Zweig, G. (2013). *Linguistic regularities in continuous space word representations*. Paper presented at the Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.
- Pennington, J., Socher, R., & Manning, C. (2014). *Glove: Global vectors for word representation*. Paper presented at the Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP).
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*.
- Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3-33): Elsevier.
- Prost, F., Thain, N., & Bolukbasi, T. (2019). *Debiasing Embeddings for Reduced Gender Bias in Text Classification*. Paper presented at the Proceedings of the First Workshop on Gender Bias in Natural Language Processing.
- Qi, P., Dozat, T., Zhang, Y., & Manning, C. D. (2018). *Universal Dependency Parsing from Scratch*. Paper presented at the Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies.
- Rind, M., & Tillinghast, L. (2008). What Is an Attributive Adjective? *Philosophy*, 83(1), 77-88. doi:10.1017/S0031819108000314
- Rudinger, R., May, C., & Van Durme, B. (2017). *Social bias in elicited natural language inferences*. Paper presented at the Proceedings of the First ACL Workshop on Ethics in Natural Language Processing.
- Swinger, N., De-Arteaga, M., Thomas Heffernan, N., Leiserson, M. D. M., & Kalai, A. T. (2019). *What are the biases in my word embedding?* Paper presented at the 2nd AAAI/ACM Conference on AI, Ethics, and Society, AIES 2019, January 27, 2019 - January 28, 2019, Honolulu, HI, United states.
- Wagner, C., Garcia, D., Jadidi, M., & Strohmaier, M. (2015). *It's a man's Wikipedia? Assessing gender inequality in an online encyclopedia*. Paper presented at the Ninth international AAAI conference on web and social media.
- Wei, W. (2019). Hierarchical division clustering framework for categorical data. *Neurocomputing*, 341, 118-135. doi:10.1016/j.neucom.2019.02.043
- Xu, R. (2015). *Clustering*: Piscataway, New Jersey : IEEE Press.
- Zhao, J., Wang, T., Yatskar, M., Cotterell, R., Ordonez, V., & Chang, K.-W. (2019). *Gender Bias in Contextualized Word Embeddings*. Paper presented at the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers).
- Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Kai-Wei, C. (2017). Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. *arXiv.org*.

- Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Kai-Wei, C. (2018). Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods. *arXiv.org*.
- Zhao, J., Zhou, Y., Li, Z., Wang, W., & Chang, K.-W. C. (2018). *Learning Gender-Neutral Word Embeddings*. Paper presented at the Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.

APPENDIX A. TOP 10 ELEMENTS FOR EACH CLUSTER

1810s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['t', 'god', 'heaven', 'lord', '@', 'doubt', 'earth', 'love', "d"]	['brazzo', 'beraldo', 'lodovico', 'hippolito', 'fool', 'beatrice', 'miss', 'astrabel', 'leonora']	['york', 'scene', 'sun', 'to- morrow', 'act', 'justice', 'david', 'carolina', 'temple']	['authority', 'form', 'spirit', 'state', 'world', 'nature', 'genius', 'three', 'states']	['april', 'michael', 'lords', 'lodo', 'jacob', 'hip', 'bell', 'porter', 'rapture']	['may', 'ought', 'will', 'might', 'shall']	['tyler', 'cooper', 'bates', 'jefferson', 'westray', 'barrett', 'oldmixon', 'milan', 'vienna']	['martin', 'carlo', 'miller', 'troth', 'hilario', 'astra', 'constancy', 'grove', 'cheerly']
Adjectives	['only', 'much', 'very', 'own', 'dear', 'little', 'subject', 'far', 'still', 'ever']	['fellow', 'lead', 'suppose', 'mean', 'do']	['early', 'pretty', 'impossible', 'clear', 'happy', 'spiritual', 'extreme', 'next', 'hard', 'free']	['many', 'same', 'whole', 'common', 'other', 'due', 'future', 'public', 'few', 'different']	['flattering', 'excellent', 'mortal', 'simple', 'pleased', 'slight', 'vulgar', 'empty', 'wretched', 'jealous']	[]	['obsolete', 'tardy', 'thirsty', "o'ertake", 'winged', 'befitting', 'ninth', 'ticklish', 'frosty', 'covetous']	['uncommon', 'abundant', 'discordant', 'dry', 'hungry', 'straight', 'marble', 'vacant', 'vigorous', 'thin']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['york', 'troth', 'to- morrow', 't', 'god', 'pacheco', 'rapture', 'horse', 'morning']	['brazzo', 'beraldo', 'lodovico', 'hippolito', 'fool', 'carlo', 'lords', 'beatrice', 'miss']	['bates', 'barrett', 'oldmixon', 'vienna', 'astragel', 'ahum', 'drazzo', 'lodovieo', 'farewel']	['league', '@', 'justice', 'drama', 'historian', 'poetry', 'fortune', 'doubt', 'man']	['april', 'michael', 'martin', 'porter', 'gentlewoman', 'poetry', 'robb', 'constancy', 'lud']	['tyler', 'cooper', 'jefferson', 'miller', 'westray', 'milan', 'judas', 'bra', 'gull']	['sun', 'carolina', 'yon', 'musick', 'march', 'bows', 'darkness', 'ray', 'lightning']	['authority', 'provision', 'state', 'christians', 'inhabitants', 'law', 'court', 'president', 'congress']
Adjectives	['many', 'early', 'pleased', 'dry', 'empty', 'wretched', 'happy', 'straight', 'glad', 'sorry']	['pretty', 'fellow', 'hard', 'dear', 'little', 'agitated', 'handsome', 'confused', 'odd', 'charming']	['frosty', 'wry', 'ampl', 'entrapp', 'mirthful', 'semicircular', 'inverse', 'ungentlemanlike', 'mahereuse', 'vitriolic']	['flattering', 'excellent', 'uncommon', 'same', 'mortal', 'simple', 'abundant', 'only', 'whole', 'much']	['slight', 'discordant', 'thirsty', 'hungry', 'vulgar', 'winged', 'ninth', 'unseen', 'marble', 'vacant']	['obsolete', 'tardy', "o'ertake", 'befitting', 'ticklish', 'covetous', 'crooked', 'worse', 'doating', 'brutal']	['clear', 'watery', 'fresh', 'deep', 'sunny', 'troubled', 'distant', 'heavy', 'broken', 'wet']	['spiritual', 'common', 'other', 'due', 'public', 'sufficient', 'different', 'proper', 'equal', 'particular']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['sir', 't', 'god']	['art', 'ha', 'de']	['april', 'michael', 'brazzo', 'tyler']	['@', 'a', 'have']	['york', 'beraldo']	['may', "d"]	['fool', 'miss']	['lodovico', 'hippolito', 'beatrice', 'leon', 'to-']

	'heaven', 'lord', 'love', 'man', 'name', 'father']	'authority', 'earth', 'form', 'honor', 'brother', 'master']	'cooper', 'bates', 'martin', 'jefferson', 'carlo']		'sun', 'duke', 'st', 'devil', 'abbess', 'fame', 'marcello']	'will', 'shall']	'scene', 'morning', 'act', 'justice', 'fortune', 'hubert', 'doubt']	morrow', 'fie', 'dost', 'theatre', 'temple']
Adjectives	['only', 'much', 'very', 'own', 'dear', 'little', 'other', 'subject', 'over', 'still']	['many', 'same', 'whole', 'happy', 'common', 'public', 'few', 'proper', 'particular', 'about']	['flattering', 'obsolete', 'uncommon', 'mortal', 'abundant', 'tardy', 'slight', 'discordant', 'dry', 'thirsty']	[]	['early', 'simple', 'clear', 'hard', 'deep', 'else', 'safe', 'bad', 'proud', 'perfect']	['do']	['fellow', 'free', 'short', 'possible', 'different', 'large', 'able', 'important', 'above', 'set']	['excellent', 'pleased', 'pretty', 'impossible', 'spiritual', 'glad', 'extreme', 'sorry', 'fond', 'next']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['lords', 'miss', 'sun', 'sir', 'rapture', 'horse', 'act', 'constancy', 'grove']	['fool', 'scene', 'troth', 'to- morrow', 'god', 'fie', 'did'st', 'historian', 'la']	['duke', 't', 'rob', 'league', 'taylor', 'poetry', 'attorney', 'heav'n', 'musick']	['thence', '@', 'man', 'count', 'king', 'wing', 'tent', 'ne', 'sion']	['st', 'de', 'th', 'der', 'le', 'des', 'ed', 'ent', 'per']	['porter', 'beat', 'morning', 'dost', 'lord', 'fame', 'earth', 'form', 'pray']	['art', 'bell', 'ha', 'heaven', 'eve', 'may', 'mar', 'tho', 'love']	['robb', 't"', 'drama', 'ah', 'huzza', '2d', 'job', 'p.', 'iii']
Adjectives	['excellent', 'mortal', 'simple', 'slight', 'dry', 'pretty', 'much', 'wretched', 'impossible', 'fellow']	['uncommon', 'abundant', 'discordant', 'thirsty', 'empty', 'jealous', 'virtuous', 'ninth', 'unseen', 'happy']	['tardy', 'hungry', 'vulgar', 'next', 'damp', 'sunny', 'due', 'peevis', 'uneasy', 'stupid']	['dear', 'aged', 'able', 'fore', 'tent', 'shed', 'over', 'rent', 'ever', 'fated']	['ing']	['flattering', 'many', 'early', 'same', 'only', 'pleased', 'whole', 'very', 'clear', 'straight']	['winged', 'thin', 'hard', 'other', 'vile', 'comely', 'willing', 'constant', 'tender', 'pure']	['giddy', 'big', 'unlucky', 'dusky', 'twixt', 'icy', 'lucky', 'odd', 'craggy', 'clumsy']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['robb', 't"', 'taylor', 'drama', 'la', 'august', 'attorney', 'heav'n', 'marr']	['st', 'de', 'don', 'th', 'ne', 'sion', 'wed', 'sue', 'der']	['fool', 'lords', 'miss', 'sir', 't', 'bell', 'ha', 'heaven', 'lord']	['porter', 'act', 'thence', 'justice', 'temple', 'authority', 'form', 'di', 'cursed']	['scene', 'sun', 'troth', 'to- morrow', 'god', 'art', 'fie', 'rapture', 'horse']	['duke', 'grove', 'league', 'did'st', 'historian', 'abbess', 'theatre', 'poetry', 'stamping']	['prihee', 'vengeance', 'enterprise', 'a', 'governor', 'serpent', 'attendants', 'conqueror', 'cousin']	['constancy', 'convent', 'perdition', 'provision', 'protection', 'state', 'ambition', 'superstition', 'caution']
Adjectives	['hungry', 'vulgar',	['dry', 'able', 'fore', 'shed',	['jealous', 'fellow', 'clear', 'sorry',	['excellent', 'mortal',	['flattering', 'many', 'early',	['abundant', 'tardy',	['uncommon', 'crooked',	['consequent', 'tent', 'consistent', 'conceited',

	'ninth', 'overbearing', 'peevis', 'uneasy', 'metaphysical', 'giddy', 'unshaken', 'big']	'ish', 'rent', 'tive', 'ing']	'silly', 'hard', 'dear', 'safe', 'bad', 'vile']	'pleased', 'whole', 'pretty', 'winged', 'impossible', 'very', 'spiritual', 'common']	'same', 'simple', 'only', 'discordant', 'much', 'empty', 'wretched']	'slight', 'thirsty', 'vacant', 'vigorous', 'damp', 'brutal', 'sunny', 'earnest']	'indulgent', 'violent', 'watery', 'damned', 'ashamed', 'troubled', 'astonished', 'licentious']	'related', 'contented', 'rational', 'considerate', 'additional', 'constitutional']
--	--	----------------------------------	--	---	--	---	---	--

1820s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['trumbull', 'canto', 'yankies', 'yankee', 'percy', 'squire', 'steer', 'ossian', 'highlands']	['salvation', 'almighty', 'bade', 'humanity', 'god', 'lord', 'providence', 'see', 'faith']	['footnote', 'town', 'eve', '1kb', 'roll', 'rail', 'sun', 'throng', 'play']	['john', 'stephen', 'walter', 'deborah', 'thomas', 'david', 'charles', 'george', 'lion']	['christians', 'boston', 'rebellion', 'prophets', 'rome', 'parliament', 'merchants', 'muse', 'scriptures']	['britain', 'state', 'powers', 'england', 'judge', 'france', 'works', 'justice', 'constitution']	['laugh', 'sung', 'corruption', 'a', 'regret', 'modesty', 'superstition', 'alarm', 'silence']	['fingal', 'fingalian', 'mac', 'jacobite', 'bute', 'o'erthrew", 'hanoverians', 'stentor', 'pinn']
Adjectives	['warlike', 'fashionable', 'potent', 'ministerial', 'pleasing', 'goodly', 'energetic', 'sympathetic', 'sour', 'copious']	['own', 'only', 'much', 'else', 'kindred', 'alone', 'set', 'dear', 'flatter', 'live']	['peaceful', 'bent', 'shut', 'clear', 'single', 'like', 'safe', 'fertile', 'thick', 'pretty']	['favorite', 'subject', 'favourite', 'individual', 'matter']	['native', 'next', 'different', 'various', 'such', 'many', 'unfit', 'whole', 'large', 'few']	['due', 'utmost', 'mental', 'religious', 'perfect', 'peculiar', 'particular', 'choice', 'professional', 'personal']	['former', 'very', 'dreadful', 'sudden', 'ordinary', 'ardent', 'wrong', 'equal', 'strange', 'feeble']	['nuisance', 'optic', 'flighty', 'antient', 'unnat'ral", 'syllogistic', 'apocalyptic', 'glib', 'damning', 'reverent']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['trumbull', 'john', 'canto', 'christians', 'boston', 'scotia', 'rebellion', 'lords', 'bishops']	['mac', "o'erthrew", 'hanoverians', 'pinn', 'popery', 'christen', 'auchmuty', 'tophet', 'booth']	['salvation', 'almighty', 'bade', 'god', 'laugh', 'providence', 'see', 'faith', 'rob']	['lenity', 'humanity', 'merchants', 'moses', 'scriptures', 'scripture', 'state', 'powers', 'homer']	['steer', 'footnote', 'highlands', 'eve', 'nile', 'barber', 'eden', '1kb', 'muses']	['fingal', 'fingalian', 'jacobite', 'stentor', "merc'ry", 'unharrass', 'starv', 'ador', 'annull']	['yankee', 'percy', 'squire', 'stephen', 'austin', 'parson', 'sam', 'deborah', 'heav']	['yankies', 'ossian', 'scot', 'sedition', 'kirk', 'bute', 'dodona', 'hessians', 'ulysses']

Adjectives	['next', 'loyal', 'politic', 'allied', 'crafty', 'new-', 'foremost', 'veteran', 'sturdy', 'academic']	["unnat'ral", 'syllogistic', 'tar-barrel', 'intenser', 'bluster', 'planetary', 'selfsame', 'welkin', 'gloomier', 'visual']	['fashionable', 'potent', 'pleasing', 'very', 'dreadful', 'sudden', 'sympathetic', 'wrong', 'strange', 'sour']	['warlike', 'native', 'former', 'energetic', 'own', 'ordinary', 'different', 'various', 'ardent', 'equal']	['goodly', 'hoary', 'bent', 'shut', 'scarlet', 'venomous', 'brazen', 'downward', 'alive', 'scared']	['apocalyptic', 'godlier', 'spir'tual', 'exc'llent', 'unmistaken', 'drachm', 'carniv'rous', 'petticoated', 'abstruser', 'cap'tal']	['silly', 'fitter', 'clever', 'undone', 'reckon', 'odd', 'foolish', 'devilish', 'ugly', 'affronted']	['ministerial', 'nuisance', 'optic', 'flighty', 'antient', 'wondrous', 'confest', 'glib', 'damning', 'reverent']
-------------------	--	---	---	---	--	---	---	---

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['christians', 'salvation', 'lords', 'rome', 'almighty', 'humanity', 'parliament', 'laugh', '1kb']	['god', 'state', 'see', 'heaven', 'may', 'death', 'states', 'nature', 'time']	['john', 'boston', 'town', 'britain', 'sun', 'charles', 'judge', 'spain', 'captain']	['yankies', 'yankee', 'steer', 'ossian', 'highlands', 'scotia', 'scot', 'sedition', 'bishops']	['footnote', 'rebellion', 'prophets', 'tories', 'eve', 'pope', 'stephen', 'bade', 'merchants']	['lord', 'powers', 'congress', 'england', 'governor', 'france', 'king', 'indians', 'faith']	['@', 'a', 'have', 'to', 'I']	['trumbull', 'canto', 'percy', 'squire', 'fingal', 'fingalian', 'mac', 'jacobite', 'o'erthrew']
Adjectives	['dreadful', 'ordinary', 'ardent', 'feeble', 'bent', 'dangerous', 'firm', 'apt', 'deadly', 'earthly']	['very', 'own', 'such', 'other', 'many', 'only', 'much', 'whole', 'like', 'few']	['native', 'sudden', 'wrong', 'strange', 'weak', 'shut', 'similar', 'clear', 'humble', 'usual']	['potent', 'ministerial', 'nuisance', 'optic', 'goodly', 'energetic', 'sympathetic', 'sour', 'hoary', 'flighty']	['warlike', 'fashionable', 'pleasing', 'helpless', 'peaceful', 'final', 'formidable', 'alarming', 'inward', 'careless']	['next', 'former', 'different', 'various', 'equal', 'single', 'large', 'important', 'latter', 'able']	[]	['copious', 'antient', 'unnat'ral', 'vicious', 'apocalyptic', 'godlier', 'oppressive', 'unfit', 'spir'tual', 'ungrateful']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['yankee', 'bishops', 'sol', 'whigs', 'chapel', 'god', 'laugh', 'forsooth', 'brush']	['salvation', 'sedition', 'prophets', 'tories', 'lenity', 'humanity', 'parliament', 'merchants', 'contempts']	['squire', 'steer', 'lords', 'town', 'eve', 'almighty', 'bade', 'dispatch', 'barber']	['lord', 'see', 'heaven', 'stood', 'aid', 'death', 'lot', 'hath', 'bless']	['christians', 'highlands', 'rebellion', 'pope', 'cooper', 'coax', 'governor', 'ensign', 'negroes']	['speed', 'reed', 'wed', 'ed', 'proceed', 'illuminated', 'blasted', 'hundred', 'breed']	['sung', 'throng', 'king', 'tilting', 'wing', 'spring', 'ring', 'wars', 'turning']	['@']
Adjectives	['warlike', 'dreadful', 'sympathetic', 'wrong', 'weak', 'helpless', 'scarlet', 'venomous', 'forlorn', 'firm']	['native', 'potent', 'ministerial', 'former', 'ordinary', 'different', 'various', 'ardent', 'equal', 'such']	['next', 'very', 'sudden', 'goodly', 'own', 'strange', 'sour', 'feeble', 'peaceful', 'hoary']	['shut', 'like', 'afraid', 'glad', 'hard', 'swell', 'handy', 'set', 'dear', 'knell']	['fashionable', 'nuisance', 'energetic', 'formidable', 'ineffable', 'vicious', 'confest', 'similar', 'melodious', 'oppressive']	['determined', 'scared', 'wicked', 'allied', 'unbounded', 'unvaried', 'variegated', 'fatigued', 'contrived', 'pleased']	['pleasing', 'damning', 'single', 'alarming', 'obliging', 'deserving', 'trifling', 'discerning', 'blustering', 'willing']	[]

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['rebellion', 'lords', 'bishops', 'prophets', 'town', 'tories', 'eve', 'pope', 'god']	['speed', 'reed', 'wed', 'ed', 'proceed', 'illuminated', 'blasted', 'breed', 'established']	['salvation', 'squire', 'sedition', 'contempts', 'providence', 'magna', 'judge', 'est', 'ensign']	['sol', 'whigs', 'almighty', 'lenity', 'laugh', 'see', 'heaven', 'grace', 'heavens']	['sung', 'king', 'tilting', 'wing', 'spring', 'ring', 'turning', 'rolling', 'lightning']	['christians', 'highlands', 'humanity', 'parliament', 'merchants', 'scriptures', 'scripture', 'divines', 'state']	['steer', 'bade', 'satan', 'forsooth', 'brush', 'roll', 'sun', 'lash', 'throng']	['yankee', 'chapel', 'barber', 'cooper', 'coax', 'm.', 'bishop', 'captain', 'cuff']
Adjectives	['warlike', 'native', 'next', 'goodly', 'peaceful', 'hoary', 'silly', 'venomous', 'firm', 'other']	['determined', 'scared', 'wicked', 'allied', 'unbounded', 'unvaried', 'variegated', 'fatigued', 'contrived', 'pleased']	['fashionable', 'potent', 'nuisance', 'very', 'energetic', 'own', 'sympathetic', 'ardent', 'such', 'sour']	['dreadful', 'wrong', 'strange', 'feeble', 'weak', 'helpless', 'bent', 'artful', 'wondrous', 'earthly']	['pleasing', 'damning', 'alarming', 'obliging', 'deserving', 'trifling', 'discerning', 'blustering', 'willing', 'insulting']	['ministerial', 'former', 'ordinary', 'different', 'various', 'equal', 'venal', 'similar', 'regular', 'venial']	['sudden', 'shut', 'flighty', 'scarlet', 'forlorn', 'wayward', 'deadly', 'clear', 'brazen', 'inward']	['glib', 'adjacent', 'dirty', 'hard', 'stubborn', 'veteran', 'sunburnt', 'lunar', 'torrid', 'billowy']

1830s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['september', 'july', 'bush', 'woods', 'island', 'valley', 'league', 'fort', 'june']	['edition', 'philadelphia', 'york', 'boston', 'district', 'pennsylvania', 'states', 'america', 'book']	['independenc e', 'heaven', 'providence', 'god', 'art', 'victory', 'soul', 'spirit', 'hast']	['campbell', 'washington', 'henry', 'porter', 'proprietor', 'pope', 'norwood', 'gertrude', '@']	['acknowledgme nt', 'may', 'agreed', 'can', 'compassion', 'writes', 'will', 'manufacture', 'production']	['wyoming', 'booksellers', 'baltimore', 'caldwell', 'northampton', 'rover', 'sharon', 'hartford', 'toryism']	['stroudsburg', 'burnside', 'hezekiah', 'susquehanna', 'sefton', 'vallette', 'uriah', 'bateman', 'zebulon']	['pierce', 'ride', 'pray', 'hark', 'mount', 'mar', 'woo', 'play', 'said']
Adjectives	['bewildered', 'rude', 'confused', 'audible', 'next', 'sunny', 'peaceful', 'agonizing', 'fertile', 'conflicting']	['many', 'strict', 'exclusive', 'local', 'absolute', 'legitimate', 'contrary', 'separate', 'respective', 'hereditary']	['much', 'further', 'secure', 'safe', 'else', 'spiritual', 'bosom', 'scorn', 'farther', 'delicacy']	['kindly', 'kindred', 'righteous', 'perfidious', 'loving', 'presumptuous', 'crafty', 'non', 'jist', 'deceased']	['doomed', 'remarkable', 'careful', 'particular', 'entire', 'wholesome', 'interesting', 'momentous', 'convinced', 'interested']	['guiltless', 'supplementary', 'oppressive', 'interminable', 'ruinous', 'surly', 'impertinent', 'immaterial', 'needful', 'inestimable']	['unparalleled', 'euphonous', 'brisker', 'traceable', 'irretraceable', 'untrammelled', '', 'kindhearted', 'unpatriotic', 'rhapsodical', 'andmy']	['dear', 'content', 'flatter', 'conceal', 'forfeit', 'save', 'forbid', 'gain', 'beat', 'mean']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
----------------------	----	----	----	----	----	----	----	----

Names	['edition', 'acknowledgeme nt', 'district', 'book', 'proprietor', 'congress', 'clerk', 'editor', 'pope']	['sharon', 'toryism', 'bateman', 'betrothment', 'forge', 'agues', 'wintermoot', 'manetto', 'englander']	['wyoming', 'philadelphia', 'booksellers', 'york', 'boston', 'baltimore', 'pennsylvania', 'september', 'states']	['rover', 'bush', 'woods', 'st', 'valley', 'hemlock', 'glade', 'fort', 'elysium']	['campbell', 'washington', 'henry', 'porter', 'norwood', 'gertrude', 'agnes', 'edward', 'watson']	['stroudsburg', 'burnside', 'sefton', 'vallette', 'austins', 'allentown', 'aranooka', 'britains', 'dorance']	['independenc e', 'heaven', 'war', '@', 'may', 'providence', 'revolution', 'divinity', 'god']	['hezekiah', 'la', 'uriah', 'silas', 'zebulon', 'v', 'harley', 'vi', 'aranooko']
Adjectives	['ruinous', 'remarkable', 'particular', 'immaterial', 'wholesome', 'interesting', 'interested', 'necessary', 'extensive', 'uniform']	['supplementa ry', 'unparalleled', 'unsophisticat ed', 'mortifying', 'unseemly', 'unobservant', 'brisker', 'traceable', 'undetermined', 'untrammelled']	['promiscuous', 'staunch', 'exhilarated', 'multitude', 'decoyed', 'individual', 'path-way', '25th', 'civilized', 'piratical']	['interminable', 'inestimable', 'audible', 'many', 'primeval', 'sunny', 'peaceful', 'vigorous', 'populous', 'retired']	['fugitive', 'feigned', 'naughty', 'crazy', 'waitin', 'afire', 'blackguard', 'lubberly', 'knowin', 'gemman']	['euphonous', 'irretraceable', 'rhapsodical', 'andmy', 'unnecessary', 'unparalleled', 'retrograde', 'retrothed', 'pre- ordained', 'forthcomingne ss']	['doomed', 'guiltless', 'oppressive', 'bewildered', 'much', 'rude', 'surly', 'confused', 'impertinent', 'careful']	['kindhearte d', 'weaponless', 'laconic', 'laughin', 'impossible', 'crusty', 'cross- examined', 'curable', 'unanswere d', 'vagrant']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['edition', 'philadelphia', 'york', 'boston', 'district', 'pennsylvania', 'september', 'independence', 'henry']	['states', 'may', 'god', 'father', 'spirit', 'house', 'world', 'mr', 'state']	['booksellers', 'caldwell', 'northampton', 'rover', 'hezekiah', 'norwood', 'susquehanna', 'sharon', 'toryism']	['baltimore', 'proprietor', 'editor', 'pope', 'gertrude', 'whigs', 'v', 'divinity', 'milton']	['washington', 'america', 'book', 'congress', 'heaven', 'war', 'indians', 'england', 'miss']	['wyoming', 'campbell', 'acknowledgme nt', 'porter', 'agnes', 'hartford', 'bush', 'hermit', 'ephrain']	['@', 'a', 'to', 'i', 'it']	['stroudsbu rg', 'burnside', 'watson', 'sefton', 'haven', 'vallette', 'mohawks', 'brandt', 'austins']
Adjectives	['rude', 'remarkable', 'careful', 'entire', 'convinced', 'interested', 'affectionate', 'splendid', 'extensive', 'uniform']	['much', 'many', 'very', 'near', 'about', 'over', 'still', 'subject', 'away', 'almost']	['guiltless', 'supplementar y', 'interminable', 'surly', 'impertinent', 'immaterial', 'unparalleled', 'unwavering', 'inestimable', 'revengeful']	['doomed', 'confused', 'desirous', 'ashamed', 'prudent', 'pleasing', 'sunny', 'romantic', 'peaceful', 'modest']	['particular', 'interesting', 'further', 'next', 'necessary', 'peculiar', 'sufficient', 'dear', 'distinguished', 'difficult']	['oppressive', 'bewildered', 'needful', 'wholesome', 'soothing', 'audible', 'sensitive', 'populous', 'irresistible', 'destructive']	['of']	['ruinous', 'momentous', 'shocked', 'peaceable', 'euphonous', 'inseparable', 'averse', 'troublesom e', 'unprinciple d', 'fervent']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
-----------------	----	----	----	----	----	----	----	----

Names	['rover', 'heaven', 'haven', 'bush', 'st', 'god', 'hemlock', 'glade', 'battle']	['l.', 's.', 'd.', 'h.', '@', 'toryism', 'v', 'ix', 'elysium']	['ned', 'agreed', 'studded', 'hied', 'laved', 'reed', 'disowned', 'blended', 'jarred']	['edition', 'acknowledgme nt', 'district', 'independence', 'proprietor', 'congress', 'indians', 'providence', 'revolution']	['booksellers', 'states', 'la', 'woods', 'island', 'valley', 'traitors', 'agues', 'volunteers']	['dining', 'gaming', 'gang', 'king', 'evening', 'lung', 'retiring', 'reading', 'thanksgiving']	['editor', 'pope', 'christians', 'doctor', 'vi', 'governor', 'victory', 'nonsense', 'continentals']	['porter', 'book', 'clerk', 'war', 'butler', 'whigs', 'miss', 'hermit']
Adjectives	['rude', 'next', 'splendid', 'sunny', 'agile', 'hearted', 'warm', 'dear', 'steep', 'narrow']	['malignant', 'sluggish', 'corrupt', 'irksome', 'grotesque', 'vague', 'poignant', 'dizzy', 'dusky', 'peevisish']	['doomed', 'bewildered', 'confused', 'convinced', 'interested', 'shocked', 'ashamed', 'unsophisticat ed', 'retired', 'qualified']	['oppressive', 'interminable', 'remarkable', 'surly', 'impertinent', 'particular', 'entire', 'immaterial', 'inestimable', 'desirous']	['twentieth', 'silvery', 'about', 'thrifty', 'thirtieth', 'thousandth', 'weighty', 'unawares', 'thirteenth', 'hundredth']	['interesting', 'unwavering', 'soothing', 'pleasing', 'agonizing', 'mortifying', 'humbling', 'conflicting', 'distinguishing', 'enterprising']	['supplementa ry', 'ruinous', 'careful', 'needful', 'wholesome', 'audible', 'peaceable', 'revengeful', 'primeval', 'patriarchal']	['guiltless', 'much', 'momentous', 'further', 'peaceful', 'modest', 'worldly', 'fertile', 'hopeless', 'morose']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['book', 'proprietor', 'divinity', 'art', 'spirit', 'superstition', 'tyranny', 'charta', 'sketch']	['@', 'butler', 'toryism', 'haven', 'may', 'miss', 'hermit', 'doctor', 'v']	['l.', 's.', 'd.', 'h.', 'porter', 'pope', 'whigs', 'la', 'island']	['ned', 'agreed', 'studded', 'hied', 'laved', 'reed', 'disowned', 'blended', 'jarred']	['booksellers', 'clerk', 'indians', 'christians', 'traitors', 'agues', 'inns', 'needles', 'mysteries']	['edition', 'acknowledgme nt', 'district', 'independence', 'states', 'congress', 'editor', 'war', 'providence']	['dining', 'gaming', 'gang', 'king', 'retiring', 'reading', 'burning', 'riding', 'swelling']	['rover', 'heaven', 'bush', 'woods', 'st', 'vi', 'valley', 'glade', 'battle']
Adjectives	['supplementary', 'interminable', 'ruinous', 'remarkable', 'careful', 'immaterial', 'needful', 'momentous', 'inestimable', 'desirous']	['guiltless', 'oppressive', 'much', 'surly', 'impertinent', 'wholesome', 'further', 'audible', 'next', 'many']	['regular', 'irregular', 'brisker', 'toilsome', 'asunder', 'loathsome', 'robust', 'paltry', 'forlorn', 'above']	['doomed', 'bewildered', 'confused', 'convinced', 'interested', 'shocked', 'ashamed', 'unsophisticated', 'retired', 'hearted']	['kindred', 'perverse', 'terse', 'graven', 'oftentimes', 'sepulchres', 'printin', 'inverse', 'sophism', 'oft- times']	['particular', 'entire', 'interesting', 'extensive', 'prudent', 'uniform', 'considerable', 'satisfactory', 'fertile', 'additional']	['unwavering', 'soothing', 'pleasing', 'agonizing', 'mortifying', 'humbling', 'conflicting', 'enterprising', 'daring', 'unwilling']	['rude', 'sunny', 'tranquil', 'steep', 'narrow', 'precipitous', 'distant', 'busy', 'boundless', 'angry']

1840s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['strike', 'hark', 'wo', 'bush', 'potash', 'kettle', 'pot', 'rob', 'medicine']	['christianity', 'god', 'lord', 'providence', 'heaven', 'mercy', '@', 'saviour', 'devil']	['berlin', 'leslie', 'harper', 'denham', 'elkington', 'carolan', 'wyndham', 'benedict', 'doughty']	['york', 'district', 'london', 'germany', 'indies', 'italy', 'england', 'america', 'january']	['charlottenbour g', 'potsdam', 'poloneus', 'hillaire', 'bourg', 'rossi', 'pyrenean', 'comtesse', 'carlsbad']	['countess', 'ida', 'norman', 'congress', 'brothers', 'clerk', 'claud', 'mary', 'charles']	['concert', 'tour', 'may', 'retribution', 'identity', 'theology', 'campaign', 'hebrew', 'dominion']	['cliff', 'street', 'office', 'park', 'continent', 'march', 'hotel', 'inn', 'fire']

Adjectives	['gorgeous', 'dear', 'huge', 'peaceful', 'swollen', 'fleshless', 'balmy', 'rude', 'luxuriant', 'grassy']	['very', 'individual', 'different', 'painful', 'distinct', 'only', 'much', 'safe', 'excited', 'utter']	['contemplative', 'vindictive', 'ingenuous', 'malignant', 'blameless', 'unprincipled', 'unreserved', 'reflective', 'heinous', 'unwise']	['many', 'several', 'next']	['debauched', 'inebriated', 'irreproachable', 'unopened', 'insufferable', 'straightforward', 'premeditated', 'irretrievable', 'ancientest', 'womanish']	['alternative', 'ignominious', 'untimely', 'attendant', 'compact', 'negative', 'paternal', 'affirmative', 'deceased', 'whispered']	['prudent', 'serious', 'unjust', 'mutual', 'crowded', 'determined', 'legal', 'advisable', 'possible', 'intelligent']	['about', 'loose', 'lifeless', 'armed', 'adjoining', 'outer', 'spacious', 'outside', 'straight', 'current']
-------------------	---	---	--	-----------------------------------	--	---	---	--

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['norman', 'office', 'christianity', 'god', 'providence', 'heaven', 'mercy', '@', 'ken']	['potsdam', 'wharton', 'hillaire', 'bourg', 'pyrenean', 'comtesse', 'hamburg', 'carlsbad', 'revoir']	['berlin', 'cliff', 'street', 'district', 'germany', 'park', 'monday', 'continent', 'march']	['ida', 'leslie', 'harper', 'claud', 'denham', 'elkington', 'carolan', 'mary', 'charles']	['strike', 'neptune', 'mariner', 'linden', 'fire', 'silence', 'bull', 'darkness', 'bird']	['charlottenbour g', 'poloneus', 'knig', 'strasse', 'kloster', 'grosses', 'nach', 'spadow', 'legans']	['york', 'congress', 'brothers', 'clerk', 'london', 'englishmen', 'siberia', 'paris', 'gentlemen']	['countess', 'count', 'seymour', 'le', 'monde', 'la', 'il', 'monsieur', 'comte']
Adjectives	['very', 'contemplative', 'prudent', 'serious', 'different', 'debauched', 'vindictive', 'unjust', 'painful', 'many']	['ancientest', 'scatheless', 'devilish', 'handsomer', 'uncarpeted', 'equestrian', 'newmade', 'crazy', 'hearted', 'unromantic']	['interminable', 'memorable', 'grassy', 'uneven', 'considerable', 'unfrequented', 'stable', 'adjoining', 'outer', 'spacious']	['homely', 'married', 'attendant', 'talented', 'cursory', 'contemporar y', 'monosyllable', 'missing', 'copious', 'extant']	['gorgeous', 'frantic', 'uncomfortable', 'tender', 'huge', 'next', 'unopened', 'glaring', 'dreadful', 'motionless']	['premeditated', 'ensurance', 'unalleviated', 'fortunehunted', 'undisguiseable', 'aproned', 'distract', 'unheardof', 'unbendable', 'everwatchful']	['individual', 'mutual', 'crowded', 'legal', 'confidential', 'alternative', 'busy', 'payable', 'related', 'nominal']	['darn', 'tarnal', 'putty', 'darned', 'afther', 'fraid', 'pesky', 'nasty', 'goin', 'drownded']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['york', 'street', 'congress', 'office', 'london', 'charles', 'strike', 'lord', 'count']	['brothers', 'clerk', 'district', 'christianity', 'germany', 'mary', 'park', 'aunt', 'edward']	['harper', 'carolan', 'charlottenbour g', 'potsdam', 'wyndham', 'poloneus', 'monday', 'vienna', 'seymour']	['countess', 'ida', 'berlin', 'norman', 'cliff', 'elkington', 'indies', 'iv', 'vi']	['god', 'may', 'being', 'people', 'time', 'state', 'states', 'father', 'soul']	['leslie', 'denham', 'benedict', 'doughty', 'beverly', 'wharton', 'digby', 'monde', 'thomson']	['claud', 'ellen', 'le', 'madame', 'v', 'hotel', 'marquis', 'monsieur', 'louis']	['@', 'a', 'will', 'to', 'one', 'are', 'that', 'i', 'for']
Adjectives	['individual', 'different', 'determined', 'possible', 'several', 'numerous', 'glad', 'dear', 'next', 'usual']	['serious', 'painful', 'mutual', 'crowded', 'legal', 'intelligent', 'agreeable', 'relative', 'correct', 'calm']	['vindictive', 'advisable', 'unprincipled', 'thoughtless', 'impotent', 'casual', 'predominant', 'pompous', 'nameless', 'imperative']	['malignant', 'worthless', 'reckless', 'frantic', 'insignificant', 'tremulous', 'glaring', 'entertaining', 'stupid', 'livid']	['very', 'many', 'only', 'much', 'about', 'like', 'ever', 'off', 'now', 'out']	['contemplative', 'debauched', 'ingenuous', 'blameless', 'unreserved', 'reflective', 'heinous', 'unwise', 'inebriated', 'impure']	['prudent', 'unjust', 'gorgeous', 'exciting', 'rational', 'lively', 'troubled', 'impressive', 'unworthy', 'vulgar']	['by']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['congress', 'district', 'christianity', 'providence', 'le', 'concert', 'al', 'citizen', 'vive']	['brothers', 'aunt', 'strike', 'god', 'lord', 'earl', 'miss', 'madame', 'monde']	['neighed', 'wed', 'bed', 'embossed', 'wrapped', 'deed', 'reed', 'burthened', 'departed']	['clerk', 'iv', '@', 'v', 'g', 'viii', 'ix', 'e', 'c.']	['cliff', 'street', 'park', 'daybreak', 'hotel', 'bob', 'bon', 'da', 'fire']	['king', 'lightning', 'being', 'whooping', 'spreading', 'praying', 'slang', 'pudding', 'ring']	['guards', 'wanderers', 'mountains', 'mountain', 'falls', 'whigs', 'flats', 'greens', 'rangers']	['countess', 'office', 'doughty', 'count', 'continent', 'march', 'scriptures', 'mercy', 'vi']
Adjectives	['contemplativ e', 'individual', 'different', 'mutual', 'legal', 'possible', 'relative', 'correct', 'rational', 'reflective']	['blameless', 'lively', 'calm', 'angry', 'dear', 'tremulous', 'ungentlemanl y', 'next', 'weak', 'likely']	['debauched', 'crowded', 'determined', 'unprincipled', 'unreserved', 'pleased', 'inebriated', 'troubled', 'disappointed', 'surprised']	['malignant', 'gorgeous', 'vulgar', 'vague', 'luxuriant', 'ashy', 'limpid', 'chivalric', 'uncouth', 'antique']	['glad', 'huge', 'straightforward', 'about', 'swollen', 'cloudy', 'balmy', 'rude', 'grassy', 'heavy']	['exciting', 'ennobling', 'glaring', 'entertaining', 'unwilling', 'provoking', 'disgusting', 'blighting', 'startling', 'appalling']	['above', 'springy', 'twofold', 'watery', 'pea-green', 'groundless', 'thousandth', 'whitish', 'fourfold', 'thousandfol d']	['very', 'prudent', 'serious', 'vindictive', 'unjust', 'painful', 'many', 'ingenuous', 'advisable', 'intelligent']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['aunt', 'strike', 'god', 'doughty', 'lord', 'miss', 'madame', 'monde', 'iv']	['countess', 'clerk', 'office', 'count', 'earl', '@', 'chapter', 'vi', 'saviour']	['cliff', 'street', 'park', 'march', 'le', 'daybreak', 'hotel', 'co', 'dieu']	['brothers', 'gentlemen', 'wanderers', 'mountains', 'tories', 'valley', 'mountain', 'hall', 'flats']	['neighed', 'wed', 'bed', 'embossed', 'wrapped', 'deed', 'reed', 'burthened', 'departed']	['congress', 'district', 'christianity', 'providence', 'scriptures', 'creator', 'concert', 'republicans', 'citizen']	['king', 'being', 'whooping', 'spreading', 'praying', 'pudding', 'ring', 'swing', 'flying']	['continent', 'indians', 'party', 'christians', 'people', 'rangers', 'rebels', 'heroes', 'heresy']
Adjectives	['blameless', 'worthless', 'lively', 'glad', 'angry', 'reckless', 'thoughtless', 'dear', 'tremulous', 'tender']	['very', 'contemplative', 'serious', 'ingenuous', 'malignant', 'advisable', 'possible', 'agreeable', 'reflective', 'heinous']	['gorgeous', 'calm', 'huge', 'straightforwar d', 'motionless', 'about', 'livid', 'cloudy', 'balmy', 'luxuriant']	['numerous', 'scanty', 'varied', 'dense', 'current', 'perpendicula r', 'fertile', 'ripe', 'parallel', 'semicircular', '']	['debauched', 'crowded', 'determined', 'unreserved', 'pleased', 'inebriated', 'troubled', 'disappointed', 'surprised', 'alarmed']	['prudent', 'individual', 'different', 'vindictive', 'unjust', 'painful', 'mutual', 'legal', 'unprincipled', 'intelligent']	['exciting', 'ennobling', 'glaring', 'entertaining', 'unwilling', 'provoking', 'disgusting', 'blighting', 'startling', 'appalling']	['many', 'several', 'characteristi c', 'civilian', 'hostile', 'commonest', 'theatric', 'impolitic', 'commoner', 'destructive']

1850s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
-----------------	----	----	----	----	----	----	----	----

Names	['sands', 'see', 'billows', 'bell', 'brook', 'dell', 'keep', 'voice', 'walk']	['folks', 'visits', 'distress', 'struggles', 'trials', 'sphere', 'masters', 'heaven', 'god']	['library', 'shop', 'watch', 'sea', 'sun', 'street', 'temple', 'house', 'year']	['al', 'redburn', 'dorchester', 'sneed', 'pawnbroker', 'gripped', 'xxi', 'johnny', 'frederic']	['charles', 'kate', 'oliver', 'alice', 'marie', 'adams', 'bobby', 'harry', 'william']	['may', 'refinement', 'improved', 'will', 'condemned', 'desire', 'possessed', 'departed', 'can']	['story', 'book', 'liverpool', 'travelers', 'series', 'church', 'morrow', 'boston', 'england']	['keller', 'haines', 'flynn', 'resorts', 'dennis', 'samaritan', 'guthrie', 'colvin', 'brattle']
Adjectives	['magnificent', , 'coarse', 'dirty', 'harsh', 'unwonted', 'about', 'painful', 'dreadful', 'exciting', 'clean']	['dear', 'immense', 'vast', 'utter', 'considerable', 'utmost', 'enormous', 'individual', 'kindred', 'trouble']	['several', 'pretty', 'extra', 'verge', 'silvery', 'flesh', 'scanty', 'lustre', 'glittering', 'faded']	['adult', 'irreconcilable', , 'menacing', 'provoking', 'husky', 'obliging', 'elated', 'tolerable', 'attentive', 'imprudent']	['much', 'else', 'wounded', 'neither', 'deceased', 'fun', 'ever', 'bid', 'badly', 'done']	['grateful', 'genial', 'interesting', 'extravagant', 'reasonable', 'determined', 'foolish', 'former', 'agreeable', 'refreshing']	['next', 'usual', 'subsequent', 'successive', 'understanding', 'earlier', 'adjacent', 'neighboring', 'material', 'medium']	['unpracticed', 'unpaid', 'sticky', 'spunky', 'knotty', 'unmercantile', , 'up-', 'stingy', 'unconquered', , 'impaired']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['shop', 'watch', 'sea', 'sun', 'street', 'atlas', 'sands', 'temple', 'hall']	['book', 'church', 'messrs', 'testament', 'school', 'court', 'bible', 'shakespeare', 'captain']	['al', 'haines', 'folks', 'vi', 'ix', 'xi', 'xvi', 'xix', 'xxi']	['keller', 'resorts', 'samaritan', 'guthrie', 'colvin', 'tremont', 'runn', 'jews', 'mccarty']	['charles', 'kate', 'redburn', 'oliver', 'alice', 'marie', 'adams', 'bobby', 'harry']	['library', 'dorchester', 'liverpool', 'travelers', 'boston', 'essex', 'england', 'manchester', 'america']	['swindlehanger', , 'freddie', 'blanchinette', 'amerel', 'ipsen', 'saletmite', 'assabeth', 'pingree', 'god-']	['story', 'visits', 'distress', 'struggles', 'series', 'trials', 'sphere', 'masters', 'heaven']
Adjectives	['magnificent', , 'refreshing', 'menacing', 'husky', 'harsh', 'about', 'attentive', 'several', 'slim', 'exciting']	['determined', 'irreconcilable', , 'adverse', 'emphatic', 'rightful', 'immediate', 'irresistible', 'opposed', 'unjust', 'influential']	['dear', 'spunky', 'stingy', 'devilish', 'spoil', 'spry', 'forewarned', 'damned', 'inter', 'coxcomb']	['unpracticed', 'unpaid', 'unfashionable', , 'galling', 'up-', , 'unconquered', 'impaired', 'downhearted', 'counterfeit', 'door-']	['foolish', 'hearty', 'afflicted', 'naughty', 'funny', 'inviting', 'pretty', 'silly', 'cordial', 'disobedient']	['ancestral', 'emancipated', 'neighboring', 'new-', 'afloat', 'lordly', 'opulent', 'suburban', 'piratical', 'domesticated']	['unmercantile', 'unjoyous', 'uncheerful', 'unappropriate', 'by-past', 'hued', 'ungladdened', 'iron-visaged', 'penetrative', 'mid-']	['grateful', 'genial', 'interesting', 'adult', 'extravagant', 'reasonable', 'former', 'agreeable', 'coarse', 'dirty']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['marie', 'simon', 'gordon', 'liverpool', 'ann', 'travelers']	['story', 'book', 'harry', 'william', 'watch', 'sea', 'sun', 'church', 'england']	['al', 'haines', 'redburn', 'oliver', 'dorchester', 'sneed']	['kate', 'adams', 'bobby', 'folks', 'visits', 'mayor', 'distress']	['@', 'a', 'will', 'i', 'can', 'one', 'are', 'an', 'have']	['keller', 'vi', 'tommy', 'manchester', 'mason', 'messrs', 'bravo']	['charles', 'alice', 'library', 'shop', 'sphere', 'boston', 'street', 'john', 'grace']	['father', 'god', 'see', 'state', 'house', 'keep', 'may', 'world', 'life']

	'hai', 'morrow', 'michael']		'pawnbroker', 'flynn', 'ix']	'struggles', 'series']		'wednesday', 'swindlehanger']		
Adjectives	['genial', 'extravagant', 'dirty', 'refreshing', 'prudent', 'tolerable', 'unwonted', 'hearty', 'attentive', 'unpleasant']	['determined', 'former', 'next', 'glad', 'possible', 'anxious', 'several', 'pleased', 'aware', 'dear']	['adult', 'irreconcilable', , 'menacing', 'husky', 'imprudent', 'hearted', 'unpaid', 'slim', 'unfashionable', , 'unlocked']	['grateful', 'reasonable', 'foolish', 'agreeable', 'magnificent', 'coarse', 'abundant', 'harsh', 'probable', 'eager']	['very', 'on', 'of']	['provoking', 'unpracticed', 'obliging', 'elated', 'pitiful', 'plausible', 'unkind', 'elderly', 'needy', 'impartial']	['interesting', 'excited', 'concerned', 'delicate', 'painful', 'acquainted', 'respectable', 'honest', 'vast', 'capable']	['much', 'about', 'sure', 'only', 'over', 'away', 'still', 'ever', 'under', 'done']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['al', 'series', 'trials', 'resorts', 'state', 'city', 'providence', 'act', 'congress']	['visits', 'travelers', 'struggles', 'sands', 'masters', 'year', 'merchants', 'savings', 'books']	['library', 't', 'i.', 'vi', 'mayor', 'ix', 'x.', 'distress', 'xi']	['book', 'harry', 'folks', 'morrow', 'father', 'heaven', 'god', 'see', 'mister']	['reed', 'felt', 'decapitated', 'creed', 'ren', 'continued', 'betrothed', 'improved', 'contested']	['king', 'dwelling', 'spring', 'mourning', 'boarding', 'fang', 'meeting', 'wedding', 'evening']	['story', 'church', 'mason', 'testament', 'school', 'court', 'bible', 'governor', 'chapter']	['pawnbroker', , 'shop', 'watch', 'sea', 'sun', 'hai', 'street', 'house', 'hall']
Adjectives	['genial', 'reasonable', 'important', 'impartial', 'sufficient', 'active', 'difficult', 'different', 'particular', 'unnecessary']	['about', 'several', 'vast', 'averse', 'thousandth', 'unawares', 'feathery', 'velvety', 'jaunty', 'later']	['grateful', 'foolish', 'dirty', 'abundant', 'unpleasant', 'useless', 'unpaid', 'honest', 'absurd', 'pitiful']	['glad', 'much', 'harsh', 'hearty', 'anxious', 'eager', 'painful', 'sure', 'dreadful', 'aware']	['determined', , 'excited', 'unpracticed', 'elated', 'concerned', 'unwonted', 'pleased', 'wretched', 'acquainted', 'hearted']	['interesting', 'refreshing', 'menacing', 'provoking', 'obliging', 'exciting', 'condescending', , 'astounding', 'unwilling', 'startling']	['adult', 'extravagant', 'irreconcilable', 'former', 'agreeable', 'magnificent', 'prudent', 'probable', 'possible', 'tolerable']	['coarse', 'next', 'husky', 'slim', 'clean', 'plump', 'brilliant', 'rude', 'vile', 'huge']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['harry',	['testament',	['al', 'story',	['sea', 'sands',	['watch',	['felt',	['king',	['series',
	'folks', 't',	'providence',	'book',	'savings',	'sun',	'decapitated',	'dwelling',	'trials',
	'pawnbroker',	'act',	'library', 'i.',	'brothers',	'heaven',	'sat', 'borne',	'spring',	'resorts',
	'shop', 'visits',	'illustration',	'mayor',	'voice',	'temple',	'continued',	'mourning',	'church',
	'vi', 'ix', 'x.']	'edition',	'struggles',	'woods', 'bats',	'house',	'ken',	'boarding',	'masters',
		'laws',	'state', 'grace']	'creek', 'corn']	'hallo', 'hall',	'betrothed',	'fang', 'meeting',	'school',
		'republic',			'tom', 'row']			'merchants',

		'revolution', 'society']				'improved', 'contested']	'wedding', 'wandering']	'district', 'scholar']
Adjectives	['foolish', 'magnificent', 'coarse', 'next', 'glad', 'much', 'delicate', 'about', 'sure', 'several']	['interesting', 'adverse', 'evident', 'sufficient', 'active', 'interested', 'different', 'emphatic', 'unnecessary', 'intimate']	['grateful', 'genial', 'adult', 'extravagant', 'reasonable', 'determined', 'irreconcilable', , 'former', 'agreeable', 'prudent']	['swampy', 'strewn', 'unawares', 'red-men', 'vexes', 'landsmen', 'stript', 'swarin', 'foemen', 'vibratory']	['dirty', 'husky', 'slim', 'clean', 'plump', 'thoughtless', 'sticky', 'knotty', 'tremendous', 'tremulous']	['excited', 'unpracticed', 'elated', 'concerned', 'unwonted', 'hearty', 'pleased', 'wretched', 'acquainted', 'hearted']	['refreshing', 'menacing', 'provoking', 'obliging', 'exciting', 'condescending', 'astounding', 'unwilling', 'startling', 'galling']	['numerous', 'scholarly', 'chiefly', 'collective', 'denotes', 'neces', 'jurymen']

1860s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['englishman', 'judge', 'providence', 'book', 'memory', 'chapter', 'bank', 'court', 'spirit']	['s', 'lip', 'jacket', 'boot', 'quarters', 'lebanon', 'pillar', 'riding', 'paradise']	['indies', 'jamaica', 'london', 'boston', 'morrow', 'cuba', '@', 'york', 'knights']	['wallingford', 'doctor', 'mary', 'jones', 'constance', 'god', 'joseph', 'tom', 'agnes']	['aggy', 'reeder', 'allens', 'deweys', 'sisyphus', 'weems', 'teller', 'digital', 'multimedia']	['will', 'may', '"d", 'might', 'lost', 'conformity', 'll', 'compromise', 'communion']	['aliens', 'willoughby', 'rick', 'ise', 'maderia', 'calcutta', 'twain', 'mecca', 'garcia']	['house', 'lane', 'cottage', 'company', 'star', 'sea', 'shop', 'island', 'hall']
Adjectives	['wretched', 'miserable', 'alone', 'much', 'possible', 'foolish', 'selfish', 'spiritual', 'faint', 'earthly']	['comfortable', 'handsome', 'about', 'hoarse', 'varying', 'lurid', 'cloudless', 'graceful', 'ghastly', 'lifeless']	['several', 'peculiar', 'particular', 'distinct', 'disciplined', 'similar', 'architectural', 'characteristic', 'verge', '28th']	['tired', 'kindly', 'rid', 'unmarried', 'badly', 'deceased', 'seldom', 'seasonable', 'gainst', 'negress']	['intimidating', 'man-flesh', 'cut- throat', 'backhanded', 'statuesque', 're- told', 'ospital', 'foreshadowing', 'unheralded', 'immoveable']	['uncomfortable', 'urgent', 'active', 'disagreeable', 'troubled', 'concerned', 'curious', 'usual', 'anxious', 'uneasy']	['sober', 'thoughtful', 'pure', 'cosy', 'enduring', 'thoughtless', 'untimely', 'distressed', 'colorless', 'husky']	['neighboring', 'armed', 'shady', 'steep', 'grassy', 'farther', 'glow', 'adjoining', 'ascending', 'margin']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['god', 'heaven', 'aliens', '@', 'knights', 'providence', 'twain', 'memory', 'will']	['house', 'maderia', 'cottage', 'hymen', 'lip', 'dreams', 'star', 'bush', 'center']	['mary', 'perkins', 'rick', 'aunt', 'aunty', 'blanche', 'massa', 'ise', 'flora']	['aggy', 'iv', 'vi', 'coroner', 'ix', 'mecca', 'chapter', 'xvi', 'xix']	['wallingford', 'doctor', 'jones', 'constance', 'joseph', 'tom', 'agnes', 'henry', 'squire']	['deweys', 'multimedia', 'cranch', 'ogres', 'jackie', 'alvarez', 'huebra', 'aguada', 'peralta']	['reeder', 'allens', 'weems', 'digital', 'pearse', 'kobbletozo', 'gog', 'magog', 'hammadhaxo']	['indies', 'jamaica', 'london', 'boston', 'morrow', 'cuba', 'york', 'england', 'ages']
Adjectives	['sober', 'uncomfortable', 'urgent', 'active', 'comfortable', 'thoughtful', 'disagreeable', 'troubled', 'pure', 'concerned']	['wretched', 'miserable', 'colorless', 'husky', 'jarring', 'warm', 'spacious', 'leathery', 'cute', 'deuced', 'beholden', 'damed']	['glad', 'sorry', 'apropos', 'bewitched', 'spunky', 'adhesiveness', 'cute', 'deuced', 'beholden', 'damed']	['intimidating', 'cosy', 'opprobrious', 'cut- throat', 'chargeable', 'inquisitorial', 'foreshadowing', 'unheralded', 'quiescent', 'immoveable']	['related', 'elderly', 'voluminous', 'accompanying', 'antiquated', 'informal', 'adept', 'heretic', 'controversial', 'dramatis']	['backhanded', 're- told', 'ospital', 'comatose', 'inner- life', 'ennuied', 'unrepelled', 're- invested', 'inter- menial', 'cortical']	['man-flesh', 'eruptive', 'morbific', 'intertwisted', 'nauseated', 'otium', 'longen', 'feverous', 'unmannered', 'extortionate']	['eventful', 'hospitable', 'intimate', 'festive', 'later', 'leisurely', 'preceding', 'subsequent', 'earlier', 'succeeding']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['god', 'house', '@', 'time', 'will', 'may', 'mr', 'state', 'friends']	['joseph', 'squire', 'abel', 'louis', 'blanche', 'jackson', 'massa']	['aliens', 'willoughby', 'perkins', 'rick', 'dean', 'ralph']	['aggy', 'constance', 'agnes', 'floyd', 'delia', 'jamaica', 'aaron', 'reeder', 'cuba']	['mary', 'henry', 'london', 'boston', 'judge', 'ages', 'america', 'memory', 'bank']	['wallingford', 'allen', 'indies', 'thompson', 'morrow', 'iv']	['doctor', 'tom', 'heaven', 'john', 'colonel', 'mrs', 'aunt', 'york', 'england']	['jones', 'adams', 'englishman', 'jacob', 'theresa', 'november', 'williams']

		'providence', 'flora']	'coroner', 'ise', 'maderia']			'bigelow', 'vi', 'knights']		'apostle', 'howard']
Adjectives	['alone', 'much', 'very', 'about', 'off', 'almost', 'look', 'still', 'even', 'away']	['sober', 'wretched', 'miserable', 'comfortable', 'thoughtful', 'troubled', 'concerned', 'vague', 'disappointed', 'absent']	['intimidating', 'cosy', 'untimely', 'colorless', 'jarring', 'outspoken', 'notable', 'minded', 'opprobrious', 'assiduous']	['man-flesh', 'husky', 'inexplicable', 'casual', 'backhanded', 'troublesome', 'reverent', 'courageous', 'inquisitorial', 'timely']	['active', 'pure', 'curious', 'usual', 'anxious', 'serious', 'likely', 'warm', 'familiar', 'slight']	['uncomfortable', 'urgent', 'enduring', 'thoughtless', 'distressed', 'advisable', 'spacious', 'flattered', 'oppressive', 'unpleasant']	['several', 'possible', 'equal', 'else', 'glad', 'safe', 'important', 'pretty', 'near', 'above']	['disagreeable', 'uneasy', 'pleasing', 'formal', 'additional', 'amused', 'exciting', 'agitated', 'varying', 'puzzled']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['house', 'knights', 'coroner', 'blanche', 'flora', 'twain', 'lane', 'cottage', 'cashier']	['doctor', 'god', 'tom', 'heaven', 'morrow', 'aunt', 'dean', 'good- morning', 'sister']	['travelling', 'king', 'riding', 'brewing', 'dancing', 'browning', 'pang', 'mong', 'sang']	['providence', 'india', 'county', 'court', 'consul', 'revelation', 'secretary', 'president', 'co']	['pictured', 'rushed', 'wed', 'indited', 'reed', 'steed', 'ned', 'ed', 'edited']	['judge', 'vi', 'memory', 'chapter', 'spirit', 'company', 'mar', 'receiver', 'story']	['aliens', 'ages', 'bank', 'libraries', 'contents', 'island', 'travels', 'cannibals', 'quarters']	['henry', 'squire', 'colonel', 'iv', '@', 'louis', 'aunty', 'ix', 'massa']
Adjectives	['husky', 'cut- throat', 'leathery', 'about', 'cheery', 'cloudless', 'keen', 'palpable', 'quiescent', 'busy']	['sober', 'thoughtful', 'thoughtless', 'untimely', 'colorless', 'alone', 'anxious', 'uneasy', 'outspoken', 'likely']	['intimidating', 'enduring', 'jarring', 'pleasing', 'startling', 'exciting', 'varying', 'surprising', 'willing', 'foreshadowing']	['urgent', 'active', 'disagreeable', 'pure', 'curious', 'usual', 'serious', 'advisable', 'several', 'notable']	['wretched', 'troubled', 'concerned', 'distressed', 'disappointed', 'minded', 'willed', 'related', 'unexplained', 'flattered']	['miserable', 'uncomfortable', 'comfortable', 'vague', 'opprobrious', 'assiduous', 'satisfactory', 'equivocal', 'discreet', 'philanthropic']	['scanty', 'dirty', 'numberless', 'sepulchres', 'vexes', 'enliven', 'velvety', 'bounden', 'unawares', 'verse- makers']	['cosy', 'statuesque', 'lurid', 'ghastly', 'lethargic', 'irrelevant', 'boyish', 'topmost', 'corrupt', 'moth- eaten']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['tom', 'squire', 'house', 'morrow', 'iv', 'aunt', 'vi', 'good-morning', 'blanche']	['god', 'henry', 'heaven', '@', 'dean', 'louis', 'aunty', 'coroner', 'ix']	['judge', 'providence', 'india', 'chapter', 'court', 'consul', 'i', 'revelation', 'secretary']	['pictured', 'rushed', 'wed', 'indited', 'reed', 'steed', 'ned', 'ed', 'edited']	['aliens', 'ages', 'libraries', 'states', 'district', 'betimes', 'tales', 'scriptures', 'revelations']	['doctor', 'colonel', 'sister', 'memory', 'will', 'may', 'spirit', 'apostle', 'co']	['travelling', 'king', 'riding', 'brewing', 'dancing', 'browning', 'bowling', 'breasting', 'ring']	['knights', 'contents', 'sea', 'island', 'cannibals', 'quarters', 'veterans', 'wives', 'dragoons']
Adjectives	['cosy', 'colorless', 'husky', 'leathery', 'about', 'fretful', 'hoarse', 'lurid', 'cheery', 'cloudless']	['sober', 'alone', 'several', 'cut- throat', 'else', 'statuesque', 'vagrant', 'anterior', 'livid', 'cursory']	['miserable', 'urgent', 'active', 'disagreeable', 'curious', 'serious', 'advisable', 'notable', 'formal', 'absent']	['wretched', 'troubled', 'concerned', 'distressed', 'outspoken', 'disappointed', 'minded', 'willed', 'related', 'unexplained']	['hors', 'quarrelsome', 'abound', 'verse- makers', 'cts', 'flourishing', 'allowd', 'principle', 'oftentimes', 'rests']	['uncomfortable', 'comfortable', 'thoughtful', 'pure', 'enduring', 'thoughtless', 'untimely', 'usual', 'anxious', 'uneasy']	['intimidating', 'jarring', 'pleasing', 'startling', 'exciting', 'varying', 'surprising', 'foreshadowing', 'alluring', 'unfading']	['clumsy', 'scanty', 'sepulchres', 'shallow', 'vexes', 'willowy', 'fellowmen', 'sea- green', 'gintlemen', 'underclothes']

1870s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['book', 'chapter', '@', 'nature', 'world', 'brothers', 'shakespeare']	['louisa', 'emma', 'ida', 'ben', 'iv', 'vii', 'viii', 'celia', 'ix']	['Sunday', 'june', 'Monday', 'march', 'plains', 'indians',	['may', 'sung', 'will', 'should', 'pathos', 'd',	['sea', 'contents', 'boy', 'bargain', 'china', 'bush',	['joanne', 'hogan', 'lilacs', 'alcott', 'carl', 'lina', 'detective',	['ride', 'tell', 'hope', 'plunge', 'dash', 'fly', 'wonder', 'walk', 'fall']	['goldy', 'mamie', 'whish', 'clarindy', 'tunnymunt', 'montebello',

	, 'europe', 'history']		'president', 'montgomery', 'york']	'writes', "ll", 'illustrated']	'moss', 'hunter', 'bears']	'bravery', 'cupid']		'coopid', 'shanghae', 'salvindoor']
Adjectives	['heavy', 'many', 'different', 'only', 'else', 'numerous', 'loving', 'further', 'ever', 'delicate']	['sayin', 'tryin', 'thinkin', 'settin', 'workin', 'damn', 'a- lookin', 'twixt', 'tellin', 'mos']	['next', 'former', 'distant', 'instant', 'memorable', 'armed', 'eventful', 'current', 'previous', 'preceding']	['polite', 'trifling', 'unable', 'pleasing', 'maternal', 'impressive', 'much', 'tired', 'unusual', 'ashamed']	['rusty', 'resplendent', 'speechless', 'superb', 'painful', 'clean', 'limp', 'stiff', 'simultaneous', 'hasty']	['overgrown', 'unlocked', 'droll', 'aforesaid', 'kindred', 'decorous', 'sunburnt', 'freckled', 'spicy', 'shiny']	['secure', 'mean', 'lose', 'endure', 'exclaimed', 'suspect', 'admit', 'tend', 'render']	['pie-crust', 'crumbly', 'restin', 'streamin', 'cretr', 'unboyish', 'exogenous', 'doggish', 'crowbackic', 'ownty']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['louisa', 'alcott', 'emma', 'ida', 'carl', 'lina', 'book', 'contents', 'ben']	['lilacs', 'sea', 'bush', 'moss', 'cinderella', 'sancho', 'plains', 'bears', 'buck']	['sunday', 'bravery', 'cupid', 'boy', 'bargain', 'june', 'nature', 'monday', 'march']	['joanne', 'hogan', 'iv', 'vii', 'viii', 'ix', 'xv', 'xvi', 'xvii']	['may', '@', 'king', 'history', 'states', 'a', 'exploration', 'society', 'bill']	['clarindy', 'tunnymunt', 'progidy', 'shanghae', 'salvindoor', 'jubal', 'houyhnhnm', 'pferd', 'kennebunkport']	['ride', 'china', 'indians', 'york', 'europe', 'spaniards', 'columbus', 'san', 'trot']	['ma', 'pooh', 'bein', 'melia', 'pa', 'mornin', 'hooray', 'injuns', 'loafin']
Adjectives	['aforesaid', 'aspiring', 'belligerent', 'uninvited', 'sulky', 'unnoticed', 'irate', 'abrupt', 'memoranda', 'verbatim']	['heavy', 'overgrown', 'unlocked', 'rusty', 'resplendent', 'superb', 'clean', 'sunburnt', 'freckled', 'spicy']	['droll', 'polite', 'trifling', 'kindred', 'speechless', 'decorous', 'unable', 'painful', 'pleasing', 'maternal']	['pie-crust', 'restin', 'unclad', 'streamin', 'afghan', 'piratical', 'featherless', 'unpropitious', 'ticklish', 'whitish']	['easy', 'possible', 'impressed', 'identical', 'interested', 'different', 'desirable', 'only', 'particular', 'valuable']	['crumbly', 'cretr', 'unboyish', 'exogenous', 'crowbackic', 'ownty', 'donty', 'unnsual', 'narry', 'dretful']	['unmolested', 'sanguinary', 'armed', 'riotous', 'enraged', 'crowded', 'civilized', 'foremost', 'circuitous', 'traditional']	['killin', 'seafarin', 'afeard', 'sayin', 'wuth', 'drownded', 'tryin', 'thinkin', 'onst', 'settin']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['contents', 'ben', 'sunday', 'ride', 'chapter', 'june', 'mary', 'march', 'indians']	['may', 'boy', 'miss', 'mrs', 'world', 'father', 'god', 'people', 'house']	['hogan', 'lilacs', 'alcott', 'carl', 'lina', 'ix', 'xv', 'xvi', 'xvii']	['emma', 'ida', 'iv', 'celia', 'detective', 'betty', 'bargain', 'jane', 'bab']	['joanne', 'louisa', 'vii', 'viii', 'thornton', 'bravery', 'goldy', 'hubbard', 'mamie']	['china', 'monday', 'hunter', 'plains', 'brothers', 'bears', 'shakespeare', 'san', 'october']	['sea', 'book', 'nature', 'president', 'john', 'york', 'king', 'sam', 'friends']	['@', 'a', 'it', 'all', 'for', 'i', 'one', 'have']
Adjectives	['heavy', 'unable', 'tired', 'unusual', 'ashamed', 'easy', 'prepared', 'interested', 'excited', 'eager']	['much', 'next', 'many', 'possible', 'only', 'off', 'ever', 'alone', 'almost', 'mean']	['overgrown', 'droll', 'decorous', 'sunburnt', 'freckled', 'spicy', 'shiny', 'pie-crust', 'limp', 'peppery']	['trifling', 'kindred', 'superb', 'pleasing', 'stiff', 'impressive', 'hasty', 'horrid', 'surprising', 'identical']	['unlocked', 'aforesaid', 'rusty', 'resplendent', 'speechless', 'maternal', 'arduous', 'crumbly', 'dishonest', 'curly']	['polite', 'painful', 'clean', 'timid', 'impressed', 'hungry', 'desirable', 'amused', 'modest', 'ample']	['afraid', 'usual', 'different', 'sorry', 'obliged', 'else', 'likely', 'able', 'former', 'further']	['by']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['sea', 'ben', 'ride', 'bravery', 'bargain', 'bush', 'moss', 'hunter', 'buck']	['contents', 'cutters', 'plains', 'states', 'pats', 'stripes', 'people', 'bill', 'tigers']	['king', 'sung', 'riding', 'hoisting', 'steering', 'drifting', 'landing', 'meeting', 'housekeeping']	['may', 'boy', '@', 'sally', 'miss', 'senor', 'world', 'brothers', 'father']	['dew', 'deed', 'mitten', 'creed', 'decayed', 'ed', 'ned', 'illustrated', 'hedged']	['detective', 'ma', 'march', 'president', 'van', 'circus', 'command', 'una', 'point']	['nature', 'history', 'lion', 'exploration', 'society', 'city', 'prince', 'dominion', 'university']	['lilacs', 'book', 'l.', 'm.', 'iv', 'vii', 'viii', 'ix', 'xv']
Adjectives	['heavy', 'droll', 'kindred', 'clean', 'shiny', 'limp', 'stiff', 'next', 'horrid', 'shaven']	['many', 'numb', 'thousandth', 'millionth', 'commonest', , 'perverse', 'spare', 'pristine', 'new-comers', 'atween']	['trifling', 'pleasing', 'surprising', 'amazing', 'loving', 'interesting', 'exciting', 'entertaining', 'uninteresting', 'willing']	['speechless', 'painful', 'much', 'easy', 'hasty', 'afraid', 'curly', 'eager', 'motherly', 'anxious']	['unlocked', 'freckled', 'tired', 'ashamed', 'unchanged', 'prepared', 'ruffled', 'impressed', 'interested', 'excited']	['polite', 'resplendent', 'superb', 'decorous', 'unable', 'arduous', 'impressive', 'unusual', 'simultaneous', 'dishonest']	['maternal', 'identical', 'unnatural', 'particular', 'unnecessary', 'intelligent', 'comical', 'unattractive', 'punctual', 'especial']	['overgrown', 'aforesaid', 'rusty', 'sunburnt', 'spicy', 'pie-crust', 'peppery', 'spoil', 'unhurt', 'gingerbread']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['lilacs', 'l.', 'm.', 'ben', 'iv', 'ix', 'xv', 'xvi', 'bravery']	['book', 'president', 'states', 'g', 'menagerie', 'bill', 'command']	['sea', 'ride', 'china', 'bush', 'moss', 'march', 'buck', 'squire', 'buttercup']	['deed', 'decayed', 'ed', 'sat', 'ned', 'closet', 'hedged']	['may', 'contents', 'vii', 'viii', 'boy', 'bargain', 'nature',	['cutters', 'plains', 'brothers', 'bears', 'pats', 'mustangs',	['sung', 'riding', 'hoisting', 'steering', 'drifting', 'landing', 'housekeeping',	['detective', 'chapter', 'history', 'exploration', 'society', 'point',

		'signorina', 'farmer']		'ken', 'blushed']	'senor', 'father']	'stripes', 'people', 'tigers']	'lightning', 'humming']	'providence', 'providences', 'university']
Adjectives	['aforesaid', 'kindred', 'sunburnt', 'spicy', 'pie-crust', 'much', 'next', 'peppery', 'many', 'hasty']	['polite', 'maternal', 'savory', 'numerous', 'former', 'ample', 'firm', 'varied', 'strict', 'immense']	['heavy', 'droll', 'rusty', 'clean', 'shiny', 'limp', 'stiff', 'horrid', 'glassy', 'upset']	['overgrown', 'unlocked', 'freckled', 'tired', 'ashamed', 'stricken', 'unchanged', 'ruffled', 'impressed', 'excited']	['resplendent', 'speechless', 'superb', 'decorous', 'unable', 'painful', 'arduous', 'impressive', 'unusual', 'easy']	['wanderin', 'new-comers', 'stocky', 'peopled']	['trifling', 'pleasing', 'surprising', 'amazing', 'loving', 'exciting', 'willing', 'aspiring', 'condescending', 'inviting']	['identical', 'interested', 'particular', 'evident', 'testy', 'interesting', 'entertaining', 'uninteresting', 'intelligent', 'untrustworthy']

1880s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['bout', 'broadway', 'tai', 'howl', 'greenwood', 'palisades', 'falls', 'ink', 'tablets']	['habberton', 'hillcrest', 'muloch', 'budgie', 'hackman', 'marfa', 'transitu', 'ound', 'zephaniah']	['@', 'providence', 'hebrew', 'will', 'scripture', 'creator', 'dictionary', 'taint', 'journals']	['helen', 'john', 'harry', 'burton', 'tom', 'alice', 'wayne', 'budge', 'toddie']	['uncle', 'lord', 'christ', 'glory', 'state', 'heaven', 'footnote', 'grandmother', 'a']	['june', 'sunday', 'sundays', 'march', 'morrow', 'eve', 'sabbath', 'saturday', 'monday']	['heavens', 'bay', 'sea', 'way', 'earth', 'mud', 'valley', 'hall', 'army']	['babies', 'york', 'children', 'massachusetts', 'democrats', 'virginia', 'louisiana', 'vienna', 'america']
Adjectives	['penetrating', 'dirty', 'bony', 'neat', 'about', 'needy', 'superb', 'bluish', 'ethereal', 'enchanted']	['crafty', 'impish', 'meditative', 'loathsome', 'baptismal', 'observant', 'unconventional', 'invulnerable', 'tasteful', 'baddy']	['impulsive', 'partial', 'disinterested', 'likely', 'much', 'uncommon', 'enough', 'considerable', 'exact', 'silly']	['fun', 'askance', 'a-comin', 'purty', 'afore', 'apropos', 'widowed', 'fixedly', 'behold', 'against']	['else', 'artistic', 'contrary', 'utter', 'particular', 'awe', 'utmost', 'momentary', 'passionate', 'choice']	['15th', '6th', '9th', '16th', '10th', '13th', '28th', '30th', '20th', '8th']	['fragrant', 'adjoining', 'inside', 'neighboring', 'depth', 'aisle', 'dimly', 'halt', 'thicket', 'precipice']	['respective', 'permanent', 'salary', 'so-called']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['hillcrest', 'budgie', 'marfa', 'ound', 'clarkson', 'hawwy', 'gloire', 'dijon', 'har']	['babies', 'children', 'morrow', 'howl', 'heavens', 'toddy', 'eve', 'bonnie', 'sea']	['commandment', 'christ', '@', 'glory', 'state', 'heaven', 'a', 'god', 'will']	['nineveh', 'vienna', 'america', 'palisades', 'hebrew', 'pagans', 'christians', 'adam', 'noah']	['june', 'york', 'sunday', 'sundays', 'massachusetts', 'march', 'broadway', 'democrats', 'virginia']	['bout', 'tai', 'hullo', 'cos', 'ar', 'nuffin', 'humph', 'darlin', 'fa']	['habberton', 'muloch', 'transitu', 'zephaniah', 'ucken', 'shong', 'harrys', 'bloomdale', 'nawndeark']	['helen', 'john', 'harry', 'burton', 'tom', 'alice', 'wayne', 'budge', 'toddie']

Adjectives	['impish', 'baddy', 'p'itty', 'a- sayin', 'hazh', 'hearted', 'cheded', 'buggy', 'brate', 'unexciting']	['penetrating', 'dirty', 'loathsome', 'bony', 'needy', 'fragrant', 'superb', 'bluish', 'ethereal', 'enchanted']	['crafty', 'impulsive', 'partial', 'disinterested', 'likely', 'meditative', 'much', 'uncommon', 'enough', 'considerable']	['baptismal', 'invulnerable', , 'unvarying', 'mutinous', 'unprejudiced', , 'portable', 'surgical', 'geometric', 'fractional', 'endearing']	['rightful', 'about', 'unfit', 'exclusive', 'previous', 'rigid', 'responsible', 'secluded', 'intense', 'subsequent']	['drowned', 'nasty', 'carryin', 'expectin', 'nat'ral', 'workin', 'chimbly', 'keerful', 'nary', 'cute']	['a-lovin', 'g'eat', 'tormed', 'ante-service', 'unexceptional', , 'fyaid', 'wockin', 'parabolic', 'netst', 'dovely']	['bed-time', 'homeless', 'irate', 'unannounced', , 'crestfallen', 'unabashed', 'wayward', 'flustered', 'trusty', 'childless']
-------------------	---	--	--	---	---	--	---	--

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['john', 'uncle', 'children', 'lord', 'state', 'god', 'will', 'england', 'way']	['alice', 'budge', 'grandmother', 'morrow', 'hayne', 'herbert', 'tai', 'louisiana', 'providence']	['habberton', 'muloch', 'hackman', 'marfa', 'transitu', 'ound', 'mayton', 'zephaniah', 'ucken']	['helen', 'tom', 'york', 'mary', 'papa', 'glory', 'massachusetts', 's', 'march', 'heaven']	['hillcrest', 'budgie', 'clarkson', 'commandment', , 'dijon', 'nineveh', 'joppa', 'mm', 'chrysostom']	['harry', 'june', 'phil', 'colonel', 'sunday', 'charles', 'christ', 'bout', 'democrats']	['babies', 'burton', 'wayne', 'toddie', 'lawrence', 'martha', 'sundays', 'jonah', 'broadway']	['@', 'a', 't', 'it', 'to', 'have']
Adjectives	['much', 'enough', 'possible', 'about', 'else', 'afraid', 'away', 'near', 'subject', 'rather']	['partial', 'uncommon', 'dirty', 'neat', 'exact', 'silly', 'stupid', 'thankful', 'patriotic', 'sensible']	['crafty', 'impish', 'disinterested', 'loathsome', 'baddy', 'rightful', 'needy', 'ethereal', 'enchanted', 'compelling']	['likely', 'usual', 'careful', 'satisfied', 'impossible', 'fit', 'necessary', 'determined', 'particular', 'anxious']	['meditative', 'bony', 'baptismal', 'observant', 'unconventional', 'invulnerable', 'tasteful', 'deluded', 'bluish', 'unvarying']	['considerable', , 'delighted', 'evident', 'enormous', 'peaceful', 'rare', 'comfortable', 'artistic', 'delicious', 'apparent']	['impulsive', 'penetrating', 'fragrant', 'superb', 'unfit', 'wistful', 'rugged', 'hoarse', 'unnatural', 'irresistible']	['on']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['tom', 'uncle', 'st', 'lord', 'bout', 'glory', 'heaven', 'grandmother', 'god']	['commandmen', 't', 'state', 'democrats', 'providence', 'india', 'dictionary', 'suspects', 'al', 'agency']	['wed', 'weed', 'enthroned', 'deed', 'reed', 'gored', 'peared', 'sen', 'illustrated']	['march', 'palisades', 'pagans', 'scripture', 'ar', 'valley', 'aden', 'journals', 'army']	['harry', 'children', 'papa', '@', 'dee', 'a', 'morrow', 'mistress', 'will']	['morning', 'ring', 'thanksgiving', 'wheeling', 'good-morning', 'wandering', 'sing', 'printing', 'king']	['budge', 's', 'hackman', 'colonel', 'mm', 'footnote', "'t", 'howl', 'toddy']	['babies', 'heavens', 'greenwood', 'falls', 'singers', 'tablets', 'lilies', 'leaves', 'grasses']

Adjectives	['loathsome', 'neat', 'about', 'bluish', 'blissful', 'fleecy', 'hoarse', 'pitiful', 'shiny', 'droll']	['partial', 'meditative', 'uncommon', 'considerable', 'unconventional', 'exact', 'usual', 'possible', 'evident', 'ethereal']	['disinterested', 'delighted', 'deluded', 'satisfied', 'rugged', 'horrificed', 'relieved', 'unprejudiced', 'determined', 'ashamed']	['dirty', 'baptismal', 'invulnerable', 'fragrant', 'verbal', 'portable', 'picturesque', 'geometric', 'vast', 'severest']	['impish', 'impulsive', 'likely', 'much', 'enough', 'silly', 'tasteful', 'stupid', 'rightful', 'thankful']	['penetrating', 'enchancing', 'compelling', 'unvarying', 'unerring', 'dazzling', 'pleasing', 'startling', 'endearing', 'ennobling']	['crafty', 'bony', 'observant', 'needy', 'unfit', 'a-sayin', 'yellowish', 'nasty', 'noisy', 'pudgy']	['flinty', 'shifty', 'toast', 'firs', 'elm-trees', 'pesterin', 'outdoors']
-------------------	--	--	--	---	---	--	---	--

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['tom', 'papa', 'glory', 'heaven', 'god', 'howl', 'heavens', 'bay', 'greenwood']	['lord', '@', 'mistress', 'will', 'bible', 'wantsh', 'creator', 'sabbath', 'way']	['commandment', 'state', 'democrats', 'providence', 'scripture', 'bishop', 'dictionary', 'journals', 'agency']	['ring', 'wheeling', 'wandering', 'sing', 'printing', 'king', 'wing', 'coting', 'bunting']	['wed', 'fright', 'enthroned', 'deed', 'gored', 'peared', 'convoyed', 'loitered', 'murdered']	['harry', 'budge', 's', 'hackman', 'colonel', 'uncle', 'children', 'st', 'mm']	['bout', 'dee', 'ee', 'tai', 'toddy', 'ar', 'nuffin', 'noffin', 'tay']	['babies', 'palisades', 'falls', 'pagans', 'singers', 'lilies', 'tokens', 'ways', 'time']
Adjectives	['enough', 'dirty', 'bony', 'neat', 'about', 'fragrant', 'bluish', 'wistful', 'fleecy', 'hoarse']	['crafty', 'impish', 'impulsive', 'meditative', 'much', 'uncommon', 'baptismal', 'usual', 'invulnerable', 'tasteful']	['partial', 'considerable', 'observant', 'unconventional', 'exact', 'possible', 'impossible', 'artistic', 'abominable', 'contrary']	['penetrating', 'enchancing', 'compelling', 'unvarying', 'unerring', 'dazzling', 'pleasing', 'startling', 'endearing', 'ennobling']	['disinterested', 'delighted', 'deluded', 'satisfied', 'rugged', 'horrificed', 'relieved', 'unprejudiced', 'drowneded', 'puzzled']	['likely', 'loathsome', 'silly', 'comfortable', 'fit', 'ashamed', 'afraid', 'uncomfortable', 'nasty', 'cowardly']	['a-sayin', 'carryin', 'fussy', 'hummin', 'trampin', 'swingin', 'a-comin', 'purty', 'tree', 'larnin']	['unawares', 'numberless', 'kindred', 'willowy', 'new-comers', 'terse', 'firs', 'oftentimes', 'four', 'elm-trees']

1890s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['isthmus', 'dynamite', 'ring', 'hell', 'rob', 'mark', 'ham', 'twain', 'galilee']	['providence', 'god', 'christ', 'christianity', 'religion', 'state', 'lord', 'progress', 'philosophy']	['jerusalem', 'march', 'column', 'york', 'africa', 'philadelphia', 'indians', 'england', 'europe']	['caesar', 'gabriel', 'houston', 'darwin', 'george', 'washington', 'stanley', 'maximilian', 'guardian']	['heinrich', 'bowery', 'broome', 'mohammedan', 'darien', 'chaucer', 'transvaal', 'hamlet', 'uri']	['city', 'broadway', 'street', 'coast', 'avenue', 'hudson', 'bay', 'hotel', 'park']	['may', 'shall', 'to-day', 'celebrated', 'will', "'d", 'said', 'heard', 'superstition', ']	['weltstein', 'uganda', 'nyanza', 'zulu', 'petition', 'bomba', 'democritus', 'heraclitus', 'nazarene']

Adjectives	['armed', 'speechless', 'gorgeous', 'wheeled', 'subterranean', 'noiseless', 'unsightly', 'luscious', 'hoarse', 'insolent']	['dear', 'much', 'spiritual', 'fundamental', 'proof', 'fun', 'its', 'responsibility', 'shock', 'delicacy']	['former', 'awful', 'else', 'lawful', 'utter', 'gradual', 'usual', 'later', 'governmental', 'incessant']	['gainst', 'purty', 'affirmative', 'reg'lar", 'layin', 'tryin', 'gwineter', 'knowed', 'ma'am", 'betrothed']	['degrading', 'marvelous', 'smokeless', 'unattached', 'aerial', 'muscular', 'scintillating', 'edible', 'alcoholic', 'laughable']	['many', 'about', 'calm', 'away', 'warm', 'grassy', 'length', 'inside', 'faint', 'steep']	['considerate', , 'responsible', 'powerful', 'impossible', 'satisfied', 'unhappy', 'wretched', 'hostile', 'intelligent', 'capable']	['communistic', , 'ulcerous', 'ampler', 'inter-continental', 'manchurian', 'gray-', 'repressive', 'mirthless', 'matutinal', 'foodless']
-------------------	---	--	---	--	---	--	---	--

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['michael', 'von', 'ut', 'revoir', 'den', 'mister', 'pa', 'git', 'chile']	['providence', 'god', 'christ', 'christianity', 'religion', 'state', 'indians', 'mohammedan', , 'republic']	['weltstein', 'uganda', 'broome', 'darien', 'nyanza', 'zulu', 'uri', 'petion', 'bomba']	['gabriel', 'houston', 'maximilian', , 'prince', 'cabano', 'guardian', 'jacob', 'shylock', 'max']	['column', 'park', 'dynamite', 'ring', 'brooks', 'hackman', 'yonder', 'destruction', 'gatling']	['jerusalem', 'march', 'city', 'york', 'africa', 'philadelphia', , 'broadway', 'bowery', 'street']	['democritus', , 'mcguiggan', 'd'aublay", 'vala', 'mell', 'puto', 'scheligan', 'nicians', 'semitized']	['caesar', 'heinrich', 'darwin', 'madison', 'chaucer', 'george', 'stanley', 'john', 'henderson']
Adjectives	['darned', 'purty', 'boun', 'orful', 'cur'us", 'anoder', 'usin', 'becomin', 'pesky', 'ligious']	['considerate', 'responsible', 'powerful', 'degrading', 'impossible', 'satisfied', 'unhappy', 'wretched', 'many', 'communistic']	['immodest', 'gainst', 'corrosive', 'gray-', 'repressive', 'matutinal', 'foodless', 'cheeked', 'unmade', 'uncountable']	['dear', 'lordly', 'undaunted', 'hateful', 'veritable', 'chatty', 'outright', 'affirmative', 'flustered', 'overjoyed']	['armed', 'wheeled', 'subterranean', 'smokeless', 'noiseless', 'unsightly', 'about', 'aerial', 'scintillating', 'edible']	['alcoholic', 'palatial', 'uniformed', 'sumptuous', 'busy', 'mountainous', , 'inhospitable', , 'roomy', 'intrenched', 'length']	['ulcerous', 'inter-continental', 'manchurian', , 'oughly', 'numerable', 'pardlike', 'anti-religion', 'under-fed', 'negroid', 'ligamentous']	['redoubtable', 'satirical', 'staunch', 'rhymed', 'escort', 'nominee', 'apothecary']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['weltstein', 'heinrich', 'uganda', 'bowery', 'broome', 'mohammedan', , 'isthmus', 'darien', 'nyanza']	['jerusalem', 'caesar', 'gabriel', 'broadway', 'darwin', 'madison', 'hamlet', 'maximilian', 'guardian']	['christ', 'religion', 'march', 'york', 'philadelphia', 'houston', 'europe', 'coast', 'india']	['@', 'a', 'have', 'will', 'I', 'it', 'one', 'we', 'no']	['stanley', 'petion', 'bismarck', 'bomba', 'shylock', 'democritus', 'nile', 'estella', 'mcguiggan']	['providence', 'christianity', 'column', 'africa', 'australia', 'china', 'japan', 'hudson', 'jacob']	['city', 'indians', 'street', 'england', 'america', 'states', 'congress', 'washington', 'evening']	['god', 'state', 'world', 'father', 'day', 'year', 'people', 'house', 'party']
Adjectives	['degrading', 'communistic', 'ampler', 'subterranean', 'smokeless', 'noiseless', 'unsightly', 'unattached']	['hostile', 'insignificant', 'gorgeous', 'marvelous', 'muscular', 'imposing', 'perplexed', 'penetrating']	['powerful', 'impossible', 'satisfied', 'intelligent', 'capable', 'desperate', 'former', 'entire']	['for']	['considerate', 'ulcerous', 'speechless', 'laborious', 'wheeled', 'inter-continental', 'witty', 'extinct']	['responsible', , 'unhappy', 'wretched', 'armed', 'tremendous', 'civilized', 'wealthy', 'polite']	['dear', 'necessary', 'possible', 'determined', 'else', 'glad', 'usual', 'later', 'due', 'mean']	['many', 'about', 'much', 'away', 'even', 'still', 'looking', 'ever', 'its', 'through']

	'aerial', 'scintillating']	'irresistible', 'courteous']	'similar', 'slight']		'manchurian', 'beneficent']	'profound', 'reasonable']		
--	-------------------------------	---------------------------------	-------------------------	--	--------------------------------	------------------------------	--	--

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['march', 'column', 'coast', 'china', 'park', 'highway', 'brooks', 'ages', 'servants']	['providence', 'religion', 'city', 'state', 'indians', 'republic', 'states', 'congress', 'information']	['prince', 'guardian', 'world', 'vestal', 'inquiry', 'president', 'professor', 'review', 'economist']	['avenue', 'isthmus', 'dynamite', '@', 'hackman', 'empire', 'chapter', 'brotherhood', 'tartar']	['ring', 'evening', 'king', 'manning', 'meeting', 'ting', 'dusting', 'praying', 'good-evening']	['street', 'bay', 'hotel', 'rob', 'earth', 'wood', 'muscle', 'twain', 'air']	['ed', 'ken', 'speed', 'trembled', 'sen', 'illustrated', 'terraced', 'weathered', 'ben']	['god', 'hamlet', 'hell', 'yonder', 'miss', 'nee', 'lord', 'ut', 'day']
Adjectives	['hostile', 'palatable', 'gainst', 'uncomfortable', 'corrosive', 'aware', 'severe', 'busy', 'damnable', 'miserable']	['considerate', 'communistic', 'necessary', 'aerial', 'entire', 'similar', 'considerable', 'reasonable', 'evident', 'actual']	['responsible', 'powerful', 'impossible', 'intelligent', 'capable', 'desperate', 'insignificant', 'tremendous', 'laborious', 'gorgeous']	['ampler', 'subterranean', 'laughable', 'aquiline', 'adroit', 'stuffy', 'vagrant', 'lurid', 'lavish', 'utter']	['degrading', 'scintillating', 'imposing', 'penetrating', 'shocking', 'striking', 'trifling', 'pleasing', 'unrelenting', 'ennobling']	['smokeless', 'slight', 'keen', 'away', 'ripe', 'rotten', 'stolid', 'gloomy', 'boundless', 'warm']	['satisfied', 'wretched', 'armed', 'civilized', 'wheeled', 'unattached', 'perplexed', 'distinguished', 'alarmed', 'obliged']	['unhappy', 'many', 'dear', 'speechless', 'noiseless', 'unsightly', 'about', 'much', 'soulless', 'calm']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['providence', 'religion', 'state', 'republic', 'isthmus', 'states', 'congress', 'information', 'park']	['god', '@', 'guardian', 'world', 'brotherhood', 'court', 'earth', 'bishop', 'ut']	['march', 'city', 'avenue', 'bay', 'hamlet', 'evening', 'police', 'hackman', 'tartar']	['indians', 'brooks', 'ages', 'servants', 'bowers', 'gentlemen', 'spies', 'vandals', 'waters']	['ring', 'king', 'manning', 'ting', 'dusting', 'praying', 'crossing', 'riding', 'bowling']	['column', 'street', 'coast', 'china', 'hotel', 'vi', 'wood', 'valley', 'col']	['ed', 'trembled', 'mought', 'terraced', 'weathered', 'drought', 'grated', 'loomed', 'seamed']	['prince', 'hell', 'yonder', 'rob', 'miss', 'father', 'grandfather', 'lord', 'ix']
Adjectives	['considerate', 'responsible', 'powerful', 'impossible', 'satisfied', 'many', 'communistic', 'hostile', 'intelligent', 'capable']	['unhappy', 'desperate', 'speechless', 'gorgeous', 'smokeless', 'noiseless', 'unsightly', 'former', 'much', 'luscious']	['about', 'slight', 'corrosive', 'aquiline', 'turbulent', 'away', 'safe', 'stuffy', 'busy', 'damnable']	['rotten', 'countless', 'canoe-men', 'mongrel', 'besides', 'carven', 'inhabit', 'outdoors', 'around', 'ranchmen']	['degrading', 'scintillating', 'imposing', 'penetrating', 'shocking', 'striking', 'trifling', 'pleasing', 'unrelenting', 'ennobling']	['hoarse', 'ripe', 'lurid', 'mountainous', 'hairy', 'rapt', 'desolate', 'shrunk', 'ablaze', 'core']	['wretched', 'armed', 'wheeled', 'unattached', 'perplexed', 'alarmed', 'uniformed', 'unmixed', 'terrified', 'incensed']	['dear', 'witty', 'shabby', 'glad', 'ashy', 'unclean', 'spare', 'careworn', 'downcast', 'akin']

1900s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['morrow', 'revolution', 'knights', 'orleans', 'eden', 'york', 'georgia', 'scotland', 'westminster']	['hospital', 'fire', 'sahara', 'wood', 'arc', 'pitcher', 'rome', 'pen', 'broadway']	['england', 'germany', 'spain', '@', 'duty', 'nature', 'independenc e', 'christianity', 'commission']	['debby', 'nelly', 'wilhelm', 'laura', 'saint', 'john', 'lord', 'satan', 'sally']	['revoitueion', , 'orpheus', 'phillinas', 'meisters', 'emerboy', 'mullein', 'hebron', 'gummage', 'barkis']	['rush', 'git', 'rob', 'drive', 'wed', 'grant', 'pardon', 'dine', 'trust']	['meister', 'confessions', 'minerva', 'watts', 'raphael', 'angelo', 'euclid', 'mignon', 'solon']	['Monday', 'october', 'june', 'Sunday', 'july', 'sleep', 'march', 'breakfast', 'three']
Adjectives	['picturesque', 'tired', 'disappointed', 'diligent', 'distasteful', 'unnecessary', 'helpful', 'comprehensiv e', 'ungrateful', 'afraid']	['mossy', 'eager', 'hurried', 'ruffled', 'exhausted', 'visible', 'shrill', 'silvery', 'strewn', 'sombre']	['many', 'much', 'capable', 'contrary', 'various', 'respective', 'involved', 'decorative', 'truly', 'rests']	['favorite', 'thinkin', 'drunk', 'nigh', 'glaring', 'leete', 'killin', 'workin', 'laughin', 'takin']	['literay', 'deomestic', 'unpropitiou s', 'crushable', 'unmasculin e', 'premonitor y', 'bulgy', 'peckish', 'pugilistic', 'unsharable']	['else', 'impossible', 'secure', 'mean', 'pretend', 'suited', 'behold', 'suspect', 'grasp', 'distinguish']	['patriarchal', 'restless', 'malignant', 'hopeless', 'unexceptionabl e', 'hurtful', 'talented', 'sisterly', 'forlorn', 'inky']	[]

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['hospital', 'confessions', 'satan', 'raphael', 'angelo', 'goethe', 'yankee', '@', 'god']	['meister', 'minerva', 'orpheus', 'watts', 'euclid', 'harper', 'mignon', 'solon', 'mullein']	['england', 'saint', 'Monday', 'germany', 'sahara', 'morrow', 'spain', 'lawrence', 'joan']	['debby', 'nelly', 'wilhelm', 'laura', 'john', 'lord', 'sally', 'michael', 'robert']	['fire', 'wood', 'piping', 'pitcher', 'glue', 'hominy', 'mar', 'rush', 'breakfast']	['revoitueion', 'phillinas', 'meisters', 'emerboy', 'gummage', 'barkis', 'sheby', 'blancs', 'letherbridge']	['tommy', 'mont', 'betsey', 'moll', 'potts', 'hullo', 'deborah', 'bess', 'jove']	['labyrinth', 'sleep', 'mercy', 'theatre', 'parthenon', 'remorse', 'myriads', 'opera', 'babel']
Adjectives	['many', 'patriarchal', 'tired', 'diligent', 'unnecessary', 'much', 'comprehensiv e', 'hurtful', 'ungrateful', 'talented']	['unexceptionabl e', 'unpropitious', 'free-', 'premonitory', 'bulgy', 'pugilistic', 'disconsolate', 'fractious', 'unsharable', 'new-']	['auspicious', 'crowded', 'convivial', 'outlying', 'spacious', 'roomy', 'palatial', 'chilly', 'commodious', 'warlike']	['glum', 'grandly', 'unconcerne d', 'humbled', 'dutiful', 'bridesmaid', , 'red-eye', 'sainted', 'rascally', 'red- handed']	['mossy', 'hurried', 'inky', 'awry', 'ruffled', 'uncultivate d', 'rustic', 'dingy', 'listless', 'grimy']	['literay', 'deomestic', 'crushable', 'unmasculine', 'peckish', 'dust-cap', 'unblown', 'wornout', 'respecttul', 'aproned']	['interestin', 'thinkin', 'leete', 'yersilf', 'pesky', 'killin', 'stummick', 'afeard', 'ornery', 'frettin']	['picturesque', 'restless', 'eager', 'disappointed', 'distasteful', 'malignant', 'hopeless', 'helpful', 'forlorn', 'comfortable']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['satan', 'michael', 'mabel', 'morrow', 'yankee']	['england', 'john', 'lord', 'fire', 'wood', 'god', 'lady', 'mrs', 'duty']	['debby', 'nelly', 'revoitueion', 'confessions', 'sahara',	['hospital', 'germany', 'spain', 'david', 'revolution',	['@', 'will', 'a', 'I', 'one', 'all', 'an', 'has', 'it']	['wilhelm', 'meister', 'minerva', 'orpheus', 'watts',	['laura', 'saint', 'sally', 'Monday', 'robert', 'lincoln',	['head', 'place', 'father', 'life', 'house', 'day', 'may', 'time', 'can']

	'tommy', 'herbert', 'lawrence', 'jacob']		'goethe', 'phillinas', 'meisters', 'harper']	'philip', 'queen', 'prince', 'maid']		'raphael', 'angelo', 'euclid', 'solon']	'rachel', 'annie', 'joan']	
Adjectives	['restless', 'unnecessary', 'helpful', 'comprehensiv e', 'forlorn', 'desolate', 'exciting', 'dingy', 'eloquent', 'rigid']	['eager', 'tired', 'afraid', 'sorry', 'satisfied', 'surprised', 'pleased', 'else', 'able', 'alone']	['literary', 'distasteful', 'deomestic', 'crushable', 'ungrateful', 'unmasculine', , 'free-', 'gratified', 'audible', 'bulgy']	['hurried', 'comfortable', , 'usual', 'aware', 'favorite', 'former', 'active', 'inevitable', 'concerned', 'enormous']	['of']	['mossy', 'patriarchal', 'diligent', 'malignant', 'unexceptionabl e', 'unpropitious', 'hurtful', 'talented', 'sisterly', 'inky']	['picturesque', 'disappointed', 'hopeless', 'thoughtful', 'exhausted', 'visible', 'generous', 'surprising', 'available', 'selfish']	['many', 'much', 'enough', 'about', 'over', 'still', 'mean', 'away', 'through', 'done']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['confessions', 'sue', 'revolution', 'nature', 'multiplication', , 'independence', 'fates', 'princess', 'india']	['knights', 'dickens', 'belles', 'lancers', 'tom', 'downs', 'stan', 'babes', 'mount']	['drew', 'wed', 'associated', 'painted', 'lined', 'graded', 'heated', 'bed', 'seed']	['hospital', 'saint', 'john', '@', 'arc', 'gladiator', 'queen', 'glue', 'hominny']	['fire', 'wood', 'st.', 'lady', 'heaven', 'sir', 'hotel', 'Sunday', 'table']	['piping', 'bowling', 'building', 'king', 'money- making', 'ring', 'reading', 'flaunting', 'wedding']	['lord', 'sally', 'morrow', 'god', 'reaper', 'prince', 'pitcher', 'maid', 'aunt']	['mont', 'duty', 'co', 'cologne', 'propinquity', 'president', 'catechism', 'mar', 'slavery']
Adjectives	['usual', 'rustic', 'active', 'complimentar y', 'energetic', 'enormous', 'attractive', 'contrite', 'frequent', 'aristocratic']	['silvery', 'mild', 'groundless', 'bare', 'frail', 'new-comers', 'watery', 'cheap', 'severest', 'weighty']	['tired', 'disappointed', , 'hurried', 'talented', 'gratified', 'ruffled', 'uncultivated', , 'exhausted', 'hearted', 'inexperience d']	['mossy', 'patriarchal', 'forlorn', 'comfortable', , 'mute', 'awry', 'mundane', 'benign', 'uppermost', 'rampant']	['aglow', 'grimy', 'shrill', 'drowsy', 'rash', 'strewn', 'pallid', 'fide', 'downcast', 'watchful']	['exciting', 'disappointing', 'dingy', 'lingering', 'aspiring', 'surprising', 'ornamenting', 'pitying', 'inviting', 'exhausting']	['eager', 'hopeless', 'much', 'helpful', 'hurtful', 'ungrateful', 'afraid', 'sisterly', 'enough', 'remorseful']	['many', 'picturesque', 'restless', 'diligent', 'distasteful', 'malignant', 'unnecessary', 'unexceptionabl e', 'unpropitious', 'comprehensive ']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['fire', 'wood', 'st.', 'heaven', 'co', 'pen', 'glue', 'sir', 'hullo']	['arc', 'cologne', 'president', 'tom', 'governor', 'commission', 'states', 'warren', 'orient']	['wed', 'associated', 'painted', 'arrival', 'lined', 'graded', 'heated', 'bed', 'seed']	['confessions', , 'dickens', 'earl', 'belles', 'lancers', 'fates', 'downs',	['hospital', 'saint', 'john', 'morrow', '@', 'god', 'knights', 'queen', 'mont']	['lord', 'sally', 'prince', 'mar', 'princess', 'rue', 'life', 'vous', 'a']	['sue', 'revolution', 'gladiator', 'duty', 'nature', 'mentor', 'propinquity', 'multiplication', 'catechism']	['piping', 'bowling', 'king', 'money- making', 'ring', 'reading', 'flaunting', 'wedding', 'sterling']

				'babes', 'despots']				
Adjectives	'mossy', 'afraid', 'desolate', 'covert', 'sorry', 'aglow', 'grimy', 'shrill', 'silvery', 'drowsy']	['patriarchal', 'premonitory', 'uppermost', 'pugilistic', 'fervent', 'rigid', 'paternal', 'belligerent', 'enormous', 'microscopic']	['tired', 'disappointed', 'hurried', 'talented', 'gratified', 'ruffled', 'aware', 'uncultivated', 'exhausted', 'hearted']	['quaint', 'new- comers', 'workaday', 'outdoors', 'everywhere', , 'indoors', 'port-holes', 'anywheres', 'semi', 'oftentimes']	['much', 'forlorn', 'comfortable', , 'enough', 'awry', 'mundane', 'benign', 'favorite', 'rampant', 'childish']	['picturesque', 'restless', 'eager', 'distasteful', 'malignant', 'hopeless', 'unpropitious', 'hurtful', 'ungrateful', 'sisterly']	['many', 'diligent', 'unnecessary', 'unexceptionabl e', 'helpful', 'comprehensive', , 'melodramatic', 'usual', 'audible', 'wasteful']	['exciting', 'disappointing', 'dingy', 'lingering', 'aspiring', 'surprising', 'ornamenting', 'pitying', 'inviting', 'exhausting']

1910s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['ascham', 'cologne', 'burgundy', 'cuba', 'titan', 'marvel', 'denver', 'investigator', 'harlem']	['italians', 'ages', 'russians', 'asia', 'girls', 'people', 'books', 'virtues', 'greeks']	['turkey', 'club', 'europe', 'harvard', 'september', 'washington', 'district', 'avenue', 'broadway']	['pettilow', 'dijon', 'melrose', 'camembert', 'ashgrove', 'lenman', 'wrenfield', 'churchman', 'papist']	['god', '@', 'law', 'fortune', 'state', 'may', 'germany', 'evolution', 'faculty']	['flood', 'bay', 'wood', 'mountain', 'lamp', 'springs', 'dawn', 'piazza', 'wall']	['grew', '"m", 'falls', 'rob', 'carry', 'look', 'win', 'escape', 'fight']	['hubert', 'granice', 'peter', 'sir', 'mr', 'miss', 'lee', 'jove', 'joseph']
Adjectives	['punctual', 'spasmodic', 'punctilious', 'sallow', 'audible', 'relentless', 'sleepless', 'unremitting', 'persistent', 'realistic']	['willing', 'about', 'many', 'various', 'numerous', 'alike', 'respective', 'so-called', 'chiefly']	['earlier', 'available', 'additional', 'preceding', 'aggregate', 'remainder', 'salary']	['unprofessional', , 'dry-mouthed', 'rid', 'dimmet', 'stoutish', 'undifferentiate d', 'unwell', 'humoured', 'extra-heavy', 'sane']	['unusual', 'much', 'underlying', 'able', 'important', 'intense', 'subtle', 'else', 'impossible', 'significant']	['distant', 'dull', 'huge', 'dingy', 'visible', 'dense', 'faint', 'damp', 'vast', 'bony']	['convinced', 'impressed', 'satisfied', 'known', 'done', 'related', 'devoted', 'stuck', 'grown', 'hunted']	['dear', 'worried', 'sorry', 'pleased', 'surprised', 'amazed', 'disappointed', , 'kindly', 'puzzled', 'stammered']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['pettilow', 'camembert', 'ranney', 'carrick', "how'do", 'eliphalet', 'slagden', 'gliddens', 'ringham']	['turkey', 'cuba', 'september', 'italians', 'district', 'avenue', 'broadway', 'harlem', 'livery']	['hubert', 'granice', 'peter', 'ascham', 'lee', 'jove', 'kate', 'denver', 'bob']	['god', 'titan', 'marvel', 'grew', 'fortune', 'daunt', 'ages', 'dredge', 'lanfear']	['suspender', 'buckle', 'buzzard', 'wood', 'lamp', 'arbor', 'springs', 'dawn', 'cashmere']	['dijon', 'melrose', 'burgundy', 'ashgrove', 'lenman', 'wrenfield', 'churchman', 'papist', 'unread']	['europe', '@', 'law', 'school', 'fittest', 'state', 'may', 'germany', 'laboratory']	['sir', 'club', 'mr', 'miss', 'cologne', 'joseph', 'harvard', 'washington', 'iii']

Adjectives	['dry-mouthed', 'extra-heavy', 'blear', 'blotchy', 'day-', 'unalive', 'gelid', 'innutritious', 'principled', 'elliptic']	['distant', 'about', 'tortuous', 'interminable', 'inaccessible', 'congested', 'vacant', 'incoming', 'spacious', 'parallel']	['dear', 'confidential', 'disgusted', 'worried', 'sorry', 'else', 'sheepish', 'pleased', 'seedy', 'surprised']	['spasmodic', 'unprofessional', 'audible', 'relentless', 'sleepless', 'unremitting', 'persistent', 'realistic', 'unrelieved', 'futile']	['sallow', 'grayish', 'huge', 'glassy', 'soda- water', 'foggy', 'comfortable', 'plump', 'ablaze', 'dingy']	['punctual', 'unescapable', 'dimmest', 'stoutish', 'undifferentiate d', 'unwell', 'senile', 'humoured', 'thousandth', 'viscous']	['punctilious', , 'willing', 'much', 'underlying', 'illegitimate', 'lawful', 'desirable', 'able', 'unsuccessful', , 'important']	['notorious', 'indefatigable', , 'explanatory', 'dedicated', 'acquainted', 'typewritten', 'laconic', 'dinner- table', 'renowned', 'widowed']
-------------------	---	--	--	--	---	---	---	---

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['granice', 'pettilow', 'dijon', 'camembert', 'papist', 'denver', 'harlem', 'ranney', 'livery']	['jove', 'joseph', 'harvard', 'september', 'iii', 'broadway', 'jury', 'attorney', 'flood']	['ascham', 'melrose', 'cologne', 'burgundy', 'ashgrove', 'lenman', 'wrenfield', 'titan', 'churchman']	['god', 'mr', 'father', 'house', 'may', 'people', 'hand', 'place', 'boy']	['hubert', 'cuba', 'italians', 'marvel', 'vi', 'vii', 'mason', 'hamilton', 'laura']	['sir', 'miss', 'europe', 'ned', 'york', 'john', 'doctor', 'lord', 'mrs']	['@', 'a', 'one', 'for', 'have', 'an', 'i', 'it', 'all']	['peter', 'turkey', 'club', 'lee', 'kate', 'washington', 'district', 'robert', 'avenue']
Adjectives	['sallow', 'audible', 'relentless', 'sleepless', 'dry- mouthed', 'realistic', 'confidential', 'meaningless', 'disgusted', 'fleeting']	['rid', 'unusual', 'worried', 'desirable', 'cheerful', 'varied', 'intense', 'subtle', 'picturesque', 'sensible']	['punctual', 'spasmodic', 'punctilious', 'unprofessional', 'unremitting', 'unrelieved', 'unescapable', 'illegitimate', 'dimmest', 'inmost']	['dear', 'much', 'about', 'many', 'else', 'known', 'almost', 'done', 'around', 'still']	['persistent', 'futile', 'underlying', 'insignificant', 'vulgar', 'threatening', 'insistent', 'urgent', 'momentary', 'plump']	['willing', 'able', 'important', 'sorry', 'impossible', 'necessary', 'difficult', 'interested', 'various', 'determined']	['of']	['distant', 'convinced', 'dull', 'huge', 'excited', 'comfortable', , 'significant', 'conscious', 'obvious', 'sufficient']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['cologne', 'titan', 'investigator', 'jury', 'unread', 'livery', 'explorer', 'vi', 'mason']	['god', 'sir', 'miss', 'lee', 'jove', 'doctor', 'good-bye', 'lord', 'father']	['district', 'law', 'school', 'providence', 'station', 'state', 'may', 'laboratory', 'evolution']	['grew', 'ed', 'promised', 'reed', 'breed', 'departed', 'seated', 'amalgamated', 'ken']	['bob', 'flood', 'buckle', 'wicks', 'hotel', 'house', 'dredge', 'bay', 'wood']	['march', 'ages', 'springs', 'sixty', 'people', 'books', 'virtues', 'limits', 'trustees']	['peter', 'turkey', 'club', 'iii', 'marvel', 'avenue', 'bosh', 'attorney', 'b.']	['ting', 'climbing', 'ironing', 'greeting', 'king', 'meeting', 'tent', 'crossing', 'sing']
Adjectives	['punctual', 'punctilious', 'audible', 'relentless', 'persistent', 'unrelieved', 'futile',	['dear', 'sleepless', 'rid', 'much', 'sordid', 'cheerful', 'unwell', 'sorry',	['unprofessional', , 'realistic', 'confidential', 'desirable', 'inert', 'equable', 'able', 'important',	['disgusted', 'convinced', 'worried', 'excited', 'varied', 'humoured', 'unmoved',	['sallow', 'dull', 'huge', 'glassy', 'plump', 'sane', 'pallid', 'florid',	['thousandth', 'centre-table', 'rare', 'vast', 'weighty', 'frontal', 'underclothes', 'bounden',	['spasmodic', 'dimmest', 'stoutish', 'grayish', 'vulgar', 'senile', 'soda-water',	['willing', 'unremitting', 'distant', 'fleeting', 'underlying', 'threatening', 'appealing',

	'unusual', 'meaningless', 'unescapable']	'about', 'else']	'incontrovertible', 'many']	'unperceived', 'furrowed', 'benumbed']	'sheepish', 'damp']	'millionth', 'canine']	'foggy', 'picturesque', 'ablaze']	'exasperating', 'dingy', 'despairing']
--	--	---------------------	--------------------------------	--	------------------------	---------------------------	---	--

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['district', 'investigator', 'jury', 'law', 'school', 'academy', 'providence', 'professor', 'state']	['ages', 'girls', 'springs', 'people', 'custodians', 'books', 'missionaries', 'children', 'misses']	['ting', 'ironing', 'greeting', 'king', 'meeting', 'crossing', 'sing', 'building', 'riding']	['bob', 'flood', 'buckle', 'dredge', 'buzzard', 'bay', 'wood', 'mountain', 'park']	['grew', 'ed', 'illustrated', 'promised', 'mought', 'reed', 'sat', 'drew', 'departed']	['club', 'titan', 'iii', 'avenue', 'attorney', 'b.', 'livery', 'explorer', 'john']	['god', 'cologne', 'marvel', 'unread', 'vi', 'vii', 'fortune', 'fittest', 'uplift']	['peter', 'turkey', 'sir', 'miss', 'lee', 'jove', 'bosh', 'doctor', 'good-bye']
Adjectives	['unprofessional', 'confidential', 'illegitimate', 'lawful', 'desirable', 'undifferentiated', 'equable', 'unsuccessful', 'important', 'incontrovertible']	['varied', 'thousandth', 'many', 'costly', 'numerous', 'riven', 'scum', 'includin', 'anywheres', 'besides']	['willing', 'unremitting', 'fleeting', 'threatening', 'exasperating', 'despairing', 'engrossing', 'deprecating', 'widening', 'astonishing']	['sallow', 'sleepless', 'distant', 'dull', 'grayish', 'huge', 'glassy', 'foggy', 'plump', 'dingy']	['disgusted', 'excited', 'humoured', 'unmoved', 'furrowed', 'benumbed', 'confused', 'enamoured', 'untrammelled', 'entranced']	['dimmest', 'stoutish', 'senile', 'soda-water', 'jovial', 'ablaze', 'nondescript', 'stagnant', 'centre-table', 'brittle']	['punctual', 'spasmodic', 'punctilious', 'audible', 'relentless', 'persistent', 'realistic', 'unrelieved', 'futile', 'unusual']	['dear', 'rid', 'worried', 'sorry', 'about', 'else', 'jaunty', 'sheepish', 'mute', 'ticklish']

1920s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['monologues', 'potpourri', 'belial', 'berkman', 'janie', 'senna', 'see-sawing', 'cyrene', 'gilead']	['sun', 'windows', 'train', 'street', 'night', 'river', 'storm', 'shadows', 'flood']	['harness', 'burning', 'honey', 'apple', 'pillars', 'match', 'mouse', 'arrow', 'neck']	['comrades', 'portraits', 'sons', 'centuries', 'greeks', 'sisters', 'england', 'men', 'people']	['jaguar', 'altitude', 'nocturne', 'scandal', 'electricity', 'skyscrapers', 'thaw', 'reveille', 'goldman']	['republic', 'poetry', 'verse', 'dream', 'secrets', 'god', '@', 'point', 'may']	['magazine', 'boy', 'iv', 'v', 'vi', 'vii', 'alexander', 'emma', 'grandpa']	['play', 'stay', 'grant', 'listen', 'sing', 'rob', 'rise', 'hang', 'drop']

Adjectives	['red-striped', 'pee-a-bed', 'gleen', 'unshapen', 'tubercular', 'unpartisan', 'lidless', 'turgid', 'unspent', 'slumbrous']	['vast', 'muffled', 'approaching', , 'outside', 'crowded', 'shadowy', 'dense', 'downward', 'towering', 'length']	['silvery', 'motionless', 'graceful', 'quivering', 'coarse', 'clenched', 'twisted', 'massive', 'wrinkled', 'greasy']	['about', 'valuable', 'numerous', 'extensive', 'respective', 'numbered', 'used', 'remaining', 'remainder', 'equipped']	['elusive', 'singed', 'ache', 'nasty', 'crinkly', 'tame', 'spiteful', 'nasal', 'avid', 'purring']	['desirable', 'favorable', 'logical', 'impossible', 'substantial', 'capable', 'satisfactory', 'remarkable', 'longing', 'concerned']	['sorry', 'pleased', 'kindly', 'retired', 'gosh', 'ago', 'frankly', 'loudly']	['else', 'secure', 'bother', 'bid', 'maintain', 'resent', 'establish', 'pretend', 'enter', 'distinguish']
-------------------	---	---	---	---	--	--	---	---

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['altitude', 'train', 'skyscrapers', 'street', 'night', 'river', 'storm', 'vi', 'flood']	['boy', 'comrades', 'emma', 'workman', 'celia', 'grandpa', 'betty', 'ling', 'christmas']	['berkman', 'cyrene', 'marchons', 'sasha', 'aldarando', 'weatherwise', 'edmundson', 'willards', 'edmondson']	['republic', 'scandal', 'electricity', '@', 'point', 'may', 'will', 'reform', 'society']	['monologues', , 'nocturne', 'potpourri', 'belial', 'veille', 'nero', 'senna', 'lebanon', 'see-sawing']	['sun', 'jaguar', 'windows', 'harness', 'alleys', 'shadows', 'burning', 'bush', 'dawn']	['poetry', 'verse', 'play', 'dream', 'secrets', 'god', 'mars', 'centuries', 'christ']	['magazine', 'iv', 'thaw', 'v', 'portraits', 'vii', 'alexander', 'goldman', 'larkin']
Adjectives	['approaching', , 'outside', 'unattended', 'fashionable', 'unannounced', , 'unoccupied', 'crowded', 'unscathed', 'mountainous', 'impassable']	['sorry', 'nasty', 'flatter', 'wry', 'pleased', 'hearty', 'excited', 'hoarse', 'anxious', 'terrified']	['pee-a-bed', 'gleen', 'unshapen', 'thoughtful', 'mettled', 'day-', 'significant', 'auroral', 'moveless', 'pulseless']	['consequent', 'desirable', 'favorable', 'unwise', 'salutary', 'substantial', 'satisfactory', 'concerned', 'undue', 'accommodating',]	['crinkly', 'red-striped', 'chested', 'tubercular', 'unpartisan', 'lidless', 'flashy', 'unspent', 'slumbrous', 'reliant']	['singed', 'silvery', 'purring', 'whiskered', 'sulphurous', 'sluggish', 'luminous', 'motionless', 'moccasin', 'succulent']	['elusive', 'ache', 'tame', 'spiteful', 'nasal', 'avid', 'vibrant', 'quiescent', 'germinal', 'interminable']	['presumptive', , 'peppery', 'unidentified', 'stalwart', 'closeted', 'retired', 'valedictory', 'staunch', 'semi-official', 'revered']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['monologues', 'nocturne', 'skyscrapers', 'potpourri', 'thaw', 'veille', 'goldman', 'larkin', 'alleys']	['republic', 'poetry', 'magazine', 'windows', 'storm', 'v', 'flood', 'christmas', 'clara']	['jaguar', 'altitude', 'belial', 'berkman', 'grandpa', 'jimmie', 'see-sawing', 'mars', 'cyrene']	['play', 'boy', 'night', 'god', 'point', 'father', 'mr', 'may', 'mrs']	['verse', 'electricity', 'iv', 'emma', 'shadows', 'joseph', 'iii', 'lizzie', 'jude']	['sun', 'dream', 'train', 'street', 'river', 'betty', 'jim', 'june', 'york']	['comrades', 'scandal', 'secrets', 'portraits', 'vi', 'vii', 'harness', 'alexander', 'workman']	['@', 'will', 'man', 'time', 'i', 'one', 'il', 'for', 'can']
Adjectives	['singed', 'crinkly', 'spiteful', 'nasal', 'avid', 'purring', 'whiskered', 'sulphurous', 'quiescent', 'sluggish']	['vast', 'desirable', 'favorable', 'capable', 'satisfactory', 'approaching', , 'remarkable', 'excited', 'peculiar', 'anxious']	['elusive', 'ache', 'nasty', 'silvery', 'red-striped', 'pee-a-bed', 'gleen', 'tame', 'vibrant', 'unshapen']	['else', 'almost', 'used', 'ago', 'also', 'before', 'through', 'known', 'around', 'along']	['motionless', 'sheer', 'shrill', 'startling', 'relaxed', 'fearful', 'exquisite', 'logical', 'picturesque', 'spoiled']	['sorry', 'impossible', 'pleased', 'concerned', 'difficult', 'determined', 'outside', 'particular', 'interesting', 'interested']	['innumerable', , 'frantic', 'obscure', 'desolate', 'muffled', 'passive', 'consequent', 'offended', 'prone', 'ambitious']	['about']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['ling', 'burning', 'fiddling', 'flying', 'piping', 'king', 'landing', 'spring', 'wheeling']	['poetry', 'verse', 'altitude', 'scandal', 'vi', 'reveille', 'workman', 'mama', 'apple']	['monologues', , 'comrades', 'skyscrapers', 'secrets', 'portraits', 'mars', 'cranes', 'sons', 'centuries']	['magazine', 'jaguar', 'iv', 'potpourri', 'v', 'vii', 'alleys', 'grandpa', 'iii']	['play', 'boy', 'dream', 'thaw', 'harness', 'god', '@', 'gem', 'honey']	['republic', 'electricity', 'canal', 'point', 'mar', 'may', 'will', 'reform', 'charity']	['wheezed', 'moved', 'chastened', 'ven', 'den', 'seated', 'ed', 'ned', 'fluffed']	['sun', 'windows', 'train', 'street', 'night', 'river', 'storm', 'shadows', 'flood']
Adjectives	['purring', 'startling', 'jarring', 'unsparing', 'exacting', 'humiliating', 'obscuring', 'enduring', 'approaching', 'unwitting']	['elusive', 'nasal', 'sulphurous', 'luminous', 'tubercular', 'indivisible', 'vast', 'motionless', 'askance', 'succulent']	['numb', 'sportmen', 'frontal', 'bounden', 'thousandth', 'overall', 'millionth', 'somewheres', 'some'eres', 'wimen']	['nasty', 'avid', 'vibrant', 'quiescent', 'sluggish', 'evanescent', 'tremulous', 'pallid', 'torpid', 'turgid']	['ache', 'sorry', 'crinkly', 'tame', 'spiteful', 'wry', 'madly', 'taut', 'hazy', 'fearful']	['germinal', 'interminable', , 'consequent', 'prone', 'favorable', 'logical', 'rustic', 'impossible', 'substantial', 'alternate']	['singd', 'whiskered', 'cheded', 'jaded', 'moccasined', 'emaciated', 'wizeded', 'warped', 'palsied', 'muffled']	['silvery', 'flatter', 'flashy', 'sheer', 'shrill', 'flaxen', 'ashen', 'greenish', 'shrunked', 'craggy']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['ling', 'burning', 'fiddling', 'flying', 'piping', 'king', 'landing', 'spring', 'wheeling']	['comrades', 'skyscrapers', 'secrets', 'portraits', 'alleys', 'mars', 'cranes', 'sons', 'centuries']	['play', 'boy', 'dream', 'night', 'thaw', 'grandpa', 'honey', 'dally', 'stay']	['poetry', 'verse', 'monologues', 'altitude', 'vi', 'vii', 'god', 'mar', 'eve']	['wheezed', 'moved', 'chastened', 'seated', 'ed', 'ned', 'fluffed', 'ted', 'reed']	['magazine', 'jaguar', 'train', 'scandal', 'street', 'iv', 'potpourri', 'v', 'reveille']	['republic', 'electricity', 'chapter', 'point', 'may', 'will', 'professor', 'reform', 'providence']	['sun', 'windows', 'river', 'storm', 'harness', 'shadows', 'flood', 'bush', 'dawn']
Adjectives	['purring', 'startling', 'jarring', 'unsparing', 'exacting', 'humiliating', 'obscuring', 'approaching', 'unwitting', 'longing']	['fertile', 'numerous', 'sportmen', 'costly', 'thousandth', 'varied', 'variegated', 'wimen', 'ques', 'atten']	['ache', 'sorry', 'nasty', 'crinkly', 'tame', 'wry', 'tearful', 'hearty', 'else', 'about']	['elusive', 'spiteful', 'avid', 'vibrant', 'sulphurous', 'quiescent', 'interminable', 'indivisible', 'vast', 'evanescent']	['singd', 'whiskered', 'cheded', 'jaded', 'moccasined', 'emaciated', 'wizeded', 'warped', 'palsied', 'relaxed']	['nasal', 'tubercular', 'askance', 'torpid', 'turgid', 'sultry', 'acrid', 'portentous', 'baser', 'goodly']	['germinal', 'consequent', 'desirable', 'favorable', 'logical', 'substantial', 'satisfactory', 'alternate', 'disciplined', 'concerned']	['silvery', 'sluggish', 'luminous', 'flatter', 'flashy', 'motionless', 'sheer', 'madly', 'pallid', 'shrill']

1930s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
-----------------	----	----	----	----	----	----	----	----

Names	['yancey', 'gee', 'granny', 'jack', 'robinson', 'ed', 'tom', 'scott', 'sam']	['giddap', 'choctaw', 'okl', 'hummm', 'mose', 'quanah', 'boudinot', 'waitie', 'mizzly']	['bush', 'tears', 'stand', 'fork', 'ice', 'hair', 'strap', 'dog', 'furniture']	['wichita', 'venable', 'sabra', 'cim', 'venables', 'isaiah', 'okla', 'cravat', 'cherokee']	['indians', 'people', 'tribes', 'soldiers', 'ladies', 'knights', 'wars', 'times', 'peoples']	['nation', '@', 'god', 'party', 'aim', 'france', 'undertaking', 'law', 'war']	['kansas', 'camp', 'territory', 'oklahoma', 'georgia', 'columbus', 'washington', 'wagon', 'california']	['states', 'bill', 'market', 'union', 'order', 'business', 'bank', 'improvement', 'association']
Adjectives	['pleased', 'startled', 'embarrassed', 'eagerly', 'grinning', 'divorced', 'ago', 'gosh', 'nod']	['up5zthe', 'toed', 'four-poster', 'hillthe', 'meager', 'unnomadic', 'forcefulthan', 'secondbestgray', 'contortionistic', 'caressingvoice']	['bony', 'shrill', 'glaring', 'shiny', 'silken', 'towering', 'massive', 'limp', 'reddish', 'slender']	['perilous', 'impractical', 'unseeing', 'awash', 'hoarse', 'imaginary', 'feverish', 'treeless', 'surprising', 'relaxed']	['about', 'numerous', 'respective', 'remaining']	['determined', 'familiar', 'inevitable', 'civilized', 'incapable', 'utmost', 'interested', 'else', 'absurd', 'formal']	['crowded', 'hurried', 'adjoining', 'neighboring', 'gone', 'headed', 'driven', 'soon', 'settled', 'farther']	['secure', 'possible', 'necessary', 'valuable', 'various', 'available', 'prospective', 'extensive', 'satisfactory', 'adequate']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['okl', 'quanah', 'boudinot', 'waitie', 'mizzly', 'isaia', 'mcalastair', 'inineteen', 'narvaez']	['nation', '@', 'proclamation', 'party', 'states', 'aim', 'mr', 'city', 'undertaking']	['people', 'emancipation', 'god', 'bible', 'testament', 'spaniard', 'socrates', 'oracle', 'word']	['robinson', 'winfield', 'scott', 'david', 'payne', 'elias', 'lincoln', 'george', 'jefferson']	['wichita', 'kansas', 'indians', 'camp', 'territory', 'oklahoma', 'tribes', 'cherokee', 'georgia']	['bush', 'cravat', 'wagon', 'wigwam', 'fork', 'ice', 'hair', 'strap', 'furniture']	['venable', 'giddap', 'venables', 'okla', 'choctaw', 'hummm', 'mose', 'comanche', 'cherokees']	['yancey', 'sabra', 'cim', 'gee', 'granny', 'jack', 'isaiah', 'ed', 'tom']
Adjectives	['up5zthe', 'hillthe', 'unnomadic', 'forcefulthan', 'secondbestgray', 'contortionistic', 'caressingvoice', 'youn', 'lostthis', 'sunbakedearth']	['secure', 'possible', 'formal', 'necessary', 'ample', 'about', 'various', 'economical', 'inadequate', 'periodic']	['perilous', 'impractical', 'determined', 'imaginary', 'feverish', 'surprising', 'conventional', 'familiar', 'sanguine', 'inevitable']	['favorite', 'taciturn', 'notorious', 'suave', 'bespectacled', 'bearded', 'youngish', 'jovial', 'mulatto', 'portly']	['afoot', 'towering', 'habitable', 'crowded', 'prehistoric', 'outside', 'consecutive', 'desultory', 'hundredth', 'desolate']	['unseeing', 'treeless', 'relaxed', 'cramped', 'four-poster', 'woolly', 'bony', 'terrific', 'ashen', 'shrill']	['awash', 'toed', 'admonitory', 'doggoned', 'mustached', 'nightmarish', 'tinny', 'acryin', 'elate', 'prideful']	['hoarse', 'worried', 'movin', 'else', 'pleased', 'laconic', 'darned', 'excited', 'frightened', 'mirthless']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['giddap', 'okla', 'okl', 'hummm', 'cravat', 'payne', 'quanah', 'boudinot', 'waitie']	['indians', 'camp', 'jack', 'territory', 'ed', 'tears', 'sam', 'david', 'lincoln']	['kansas', 'gee', 'granny', 'robinson', 'georgia', 'scott', 'parker', 'jefferson', 'davis']	['people', 'stand', 'god', 'states', 'mr', 'house', 'may', 'man', 'times']	['wichita', 'venable', 'cim', 'venables', 'isaiah', 'choctaw', 'mose', 'cherokee', 'winfield']	['tom', 'nation', 'washington', 'george', 'john', 'party', 'hair', 'joe', 'dog']	['@', 'a', 'i', 'one', 'have', 'it', 'an', 'all', 'of']	['yancey', 'sabra', 'bush', 'oklahoma', 'tribes', 'columbus', 'proclamation', 'floyd', 'howdy']

Adjectives	['perilous', 'up5zthe', 'feverish', 'cramped', 'hillthe', 'meager', 'unnomadic', 'unprepared', 'forcefulthan', 'secondbestgray']	['determined', 'familiar', 'inevitable', 'secure', 'worried', 'interested', 'pleased', 'formal', 'excited', 'valuable']	['surprising', 'relaxed', 'civilized', 'terrific', 'varied', 'delightful', 'utmost', 'absurd', 'impatient', 'accustomed']	['possible', 'about', 'afraid', 'even', 'almost', 'away', 'along', 'mean', 'gone', 'ever']	['impractical', 'unseeing', 'awash', 'toed', 'treeless', 'four- poster', 'woolly', 'sanguine', 'incongruous', 'ashen']	['else', 'necessary', 'various', 'impossible', 'concerned', 'outside', 'subject', 'convinced', 'finished', 'suddenly']	['hoarse', 'imaginary', 'conventional', , 'bony', 'incapable', 'solitary', 'shrill', 'rude', 'distressed', 'uneasy']
-------------------	---	--	--	--	---	---	--

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['bush', 'camp', 'jack', 'tom', 'cravat', 'tears', 'stand', 'lak', 'wagon']	['nation', 'emancipation', 'proclamation', 'party', 'states', 'city', 'law', 'testament', 'war']	['ed', 'ven', 'freed', 'ted', 'den', 'ken', 'bended', 'ignited', 'seven']	['gee', 'granny', 'john', 'smith', '@', 'god', 'howdy', 'aim', 'wigwam']	['people', 'tribes', 'fork', 'parlors', 'coast', 'squatters', 'sooners', 'plains', 'tres']	['ain', 'avenue', 'dodge', 'ci', 'coroner', 'oracle', 'pottage', 'faro', 'cologne']	['territory', 'deputy', 'marshal', 'ame', 'noth', 'bible', 'furniture', 'mar', 'quince']	['ning', 'undertaking', 'king', 'rooming', 'browning', 'building', 'ng', 'boating', 'sleeping']
Adjectives	['awash', 'treeless', 'shrill', 'movin', 'leathery', 'glassy', 'shiny', 'silken', 'florid', 'glib']	['impractical', 'conventional', 'solitary', 'secure', 'formal', 'necessary', 'furtive', 'various', 'characteristic', 'economical']	['toed', 'relaxed', 'cramped', 'unprepared', 'varied', 'worried', 'distressed', 'abashed', 'accustomed', 'doggoned']	['hoarse', 'woolly', 'sanguine', 'ashen', 'delightful', 'rude', 'else', 'uneasy', 'absurd', 'mournful']	['hundredyears', 'hundredth', 'thousanth', 'unawares', 'indoors', 'firs', 'millionth', 'quare', 'thousan', 'stem']	['four-poster', 'bony', 'tireless', 'hospitable', 'cumbersome', , 'nightmarish', 'vibrant', 'lugubrious', 'grotesque', 'languid']	['perilous', 'determined', 'imaginary', 'feverish', 'familiar', 'inevitable', 'meager', 'civilized', 'incapable', 'terrific']	['unseeing', 'surprising', 'exhilarating', 'blistering', 'menacing', 'glaring', 'despairing', 'towering', 'astounding', 'terrifying']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['people', 'tribes', 'furniture', 'parlors', 'coast', 'squatters', 'sooners', 'plains', 'tres']	['camp', '@', 'marshal', 'avenue', 'saloon', 'store', 'whither', 'ci', 'pottage']	['proclamation', , 'party', 'deputy', 'ame', 'city', 'testament', 'governor', 'grande', 'bill']	['bush', 'jack', 'tom', 'cravat', 'tears', 'lak', 'wagon', 'fork', 'ice']	['territory', 'nation', 'emancipation', 'states', 'aim', 'undertaking', 'law', 'war', 'al']	['ed', 'ning', 'king', 'rooming', 'browning', 'ng', 'ted', 'boating', 'sleeping']	['noth', 'bible', 'abou', 'wilderness', 'tal', 'oracle', 'word', 'ut', 'arc']	['gee', 'granny', 'stand', 'john', 'smith', 'ain', 'god', 'howdy', 'wigwam']

Adjectives	['hundredyears', 'peopled', 'larly', 'thousandth', 'varicose', 'evens', 'indoors', 'firs', 'bulk', 'thousan']	['awash', 'four- poster', 'meager', 'hospitable', 'pungent', 'cumbersome', 'nightmarish', 'longish', 'lugubrious', 'scant']	['conventional', , 'interested', 'admonitory', 'formal', 'offended', 'courtly', 'grandiloquent', , 'favorite', 'addicted', 'sartorial']	['treeless', 'woolly', 'bony', 'ashen', 'shrill', 'movin', 'flimsy', 'tinny', 'ornate', 'grotesque']	['impractical', 'varied', 'secure', 'possible', 'accommodating', , 'necessary', 'valuable', 'various', 'economical', 'inadequate']	['unseeing', 'toed', 'relaxed', 'cramped', 'distressed', 'abashed', 'restless', 'doggoned', 'darned', 'mustached']	['perilous', 'determined', 'imaginary', 'feverish', 'surprising', 'familiar', 'sanguine', 'inevitable', 'civilized', 'incapable']	['hoarse', 'worried', 'rude', 'else', 'impatient', 'mournful', 'pleased', 'a- cryin', 'mirthless', 'about']
-------------------	---	--	--	---	--	---	--	---

1940s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['day', 'bureau', 'washington', 'harbor', 'december', 'york', 'september', 'monday', 'august']	['stechers', 'sonderheim', 'goddamnness', 'gallonsand', 'dumm', 'underthat', 'talty', "ass't", 'pthuh']	['money', 'states', 'scrap', 'jews', 'associates', 'times', 'trades', 'offices', 'tickets']	['show', 'die', 'finish', 'borrow', 'sing', 'hell', 'fight', 'drive', 'slip']	['order', 'god', 'department', 'press', '@', 'europe', 'book', 'india', 'man']	['decoration', 'payne', 'mohawk', 'hooray', 'stecher', 'lottie', 'flossie', 'newport', 'hercules']	['gurlie', 'pilot', 'joe', 'president', 'caldwell', 'director', 'daddy', 'douglas', 'alice']	['bar', 'horse', 'water', 'face', 'wheel', 'mule', 'milk', 'hay', 'fire']
Adjectives	['stationed', 'hurried', 'adjoining', 'gone', 'parked', 'newly']	['rottenthrough', 'nothinghappene d', 'heraldic', 'indianslived', 'disobedient', 'natured', 'underhand', 'ex- collegiate', 'cheeked', 'larval']	['various', 'alone', 'numerous', 'respective', 'chiefly', 'equipped', 'so- called', 'together', 'remaining', 'ranging']	['please', 'preserve', 'announce', 'handed', 'remove', 'refuse', 'survive', 'grab', 'tend', 'disappear']	['interested', 'possible', 'slight', 'else', 'indispensabl e', 'necessary', 'entire', 'likely', 'formal', 'impossible']	['prone', 'scared', 'methodical', 'thorough', 'unmixed', 'amused', 'amusing', 'familiar', 'dignified', 'steamy']	['embarrasse d', 'sorry', 'surprised', 'kindly', 'puzzled', 'startled', 'delighted', 'annoyed', 'ago', 'gosh']	['blond', 'coarse', 'packed', 'taut', 'slender', 'away', 'dull', 'backward', 'thick', 'faint']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['sonderheim', 'goddamnness', 'gallonsand', 'dumm', 'underthat', 'talty', "ass't", 'pthuh', 'storch']	['payne', 'president', 'washington', 'caldwell', 'director', 'douglas', 'roosevelt', 'york', 'smith']	['decoration', 'bar', 'horse', 'whang', 'derby', 'water', 'hoe', 'dab', 'wheel']	['stechers', 'gunnar', 'amberggris', 'perce', 'wynnewood', 'grantland', 'brittain', 'fagan', 'stasia']	['day', 'pilot', 'mohawk', 'harbor', 'europe', 'december', 'newport', 'hercules', 'september']	['gurlie', 'joe', 'god', 'hooray', 'stecher', 'lottie', 'flossie', 'daddy', 'alice']	['show', 'man', 'shakespeare', , 'face', 'jesus', 'curiosity', 'americans', 'betrayal', 'americanism ']	['money', 'order', 'bureau', 'department', , 'states', 'press', '@', 'scrap', 'office']
Adjectives	['rottenthrough', 'indianslived', 'ex- collegiate', 'accidentaleconom ic', 'surroundthe', 'unruled', 'certainknowledge', , 'toughest']	['jovial', 'spectacled', 'crusty', 'courtly', 'suave', 'shrewd', 'affable', 'hardworking']	['steamy', 'muggy', 'fireproof', 'mottled', 'blond', 'lazy', 'petrified', 'coarse', 'yellowish', 'uneven']	['nothinghappen ed', 'heraldic', 'disobedient', 'natured', 'constipated', 'underhand', 'cheeked', 'larval', 'withhis', 'saleable']	['desolate', 'away', 'conscript', 'congested', 'divisional', 'uncounted', 'unnoticed', 'adrift']	['scared', 'amused', 'embarrassed', , 'comfortable', , 'sorry', 'cute', 'else', 'excited']	['interested', 'prone', 'methodical', 'unmixed', 'amusing', 'familiar', 'dignified', 'distinct',	['thorough', 'possible', 'various', 'necessary', 'entire', 'qualified', 'non-union', 'unfair',

	'snouted', 'differentwhen']	'facetious', 'personable']			'afloat', 'unbroken']	'surprised', 'worried']	'slight', 'formless']	'bona', 'wasteful']
--	--------------------------------	-------------------------------	--	--	--------------------------	----------------------------	--------------------------	------------------------

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['money', 'order', 'god', 'president', 'states', 'washington', 'europe', 'show', 'york']	['stechers', 'sonderheim', 'goddammess', 'hooray', 'stecher', 'newport', 'gallonsand', 'dumm', 'raleigh']	['pilot', 'bureau', 'director', 'harbor', 'december', 'smith', 'september', 'monday', 'august']	['day', '@', 'man', 'war', 'may', 'people', 'will', 'time', 'i']	['payne', 'mohawk', 'lottie', 'flossie', 'hercules', 'wesson', 'nana', 'whang', 'gunnar']	['decoration', 'caldwell', 'daddy', 'herald', 'sarah', 'vermont', 'katherine', 'devery', 'blake']	['gurlie', 'douglas', 'alice', 'wallace', 'associates', 'baltimore', 'printing', 'carl', 'harris']	['joe', 'department', 'press', 'bar', 'roosevelt', 'horse', 'book', 'doctor', 'center']
Adjectives	['possible', 'else', 'necessary', 'alone', 'less', 'ago', 'along', 'trying', 'gone', 'suddenly']	['prone', 'rotentthrough', 'nothinghappene d', 'indianslived', 'subdued', 'thrilled', 'ex- collegiate', 'cheeked', 'uneven', 'larval']	['scared', 'slight', 'comfortable', 'excited', 'inclined', 'weak', 'packed', 'frightened', 'impressed', 'formal']	['away', 'even', 'over', 'done', 'more', 'too', 'through']	['methodical', , 'unmixed', 'steamy', 'muggy', 'heraldic', 'fireproof', 'formless', 'tame', 'disobedient', 'mottled']	['dignified', 'cute', 'lazy', 'indispensabl e', 'taut', 'qualified', 'unfair', 'hearty', 'reminiscent', 'fascinated']	['thorough', 'amused', 'amusing', 'embarrassed', , 'distinct', 'relaxed', 'blond', 'differ', 'coarse', 'healthy']	['interested', , 'familiar', 'sorry', 'surprised', 'worried', 'various', 'entire', 'likely', 'impossible', , 'concerned']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['center', 'times', 'trades', 'mark', 'offices', 'island', 'inks', 'mach', 'milk']	['pilot', 'bar', 'horse', 'scrap', 'st.', 'boss', 'bridge', 'water', 'street']	['building', 'printing', 'sing', 'repairing', 'flying', 'king', 'ring', 'ong', 'riding']	['decoration', 'order', 'department', 'president', 'states', 'press', 'city', 'associates', 'police']	['gran', 'ed', 'embalmed', 'den', 'tailed', 'weed', 'ted', 'seven', 'staked']	['bureau', 'director', 'office', 'derby', 'jury', 'inspectors', 'chase', 'nonsense', 'devil']	['joe', 'daddy', 'harbor', 'smith', 'j.', 'd.', 'c.', 'august', 'postmaster']	['day', 'money', 'god', '@', 'show', 'whang', 'doctor', 'sir', 'man']
Adjectives	['rare', 'thousandth', 'unawares', 'portly', 'weighty', 'groundless', 'neces', 'anywheres', 'millionth', 'marketable']	['steamy', 'slight', 'blond', 'coarse', 'mellow', 'straightforward', , 'slender', 'away', 'dull', 'oily']	['amusing', 'compelling', 'jarring', 'painstaking', 'damaging', 'unmoving', 'terrifying', 'unseeing',	['prone', 'methodical', 'possible', 'constipated', 'various', 'necessary', 'entire', 'civilized',	['scared', 'unmixed', 'amused', 'embarrassed', , 'dignified', 'relaxed', 'mottled', 'excited',	['interested', 'thorough', 'familiar', 'amused', 'distinct', 'heraldic', 'formless', 'disobedient', 'differ', 'healthy',	['muggy', 'fireproof', 'cute', 'lazy', 'stylish', 'yellowish', 'unscrupulou s', 'bona', 'grimy', 'swarthy']	['tame', 'comfortabl e', 'sorry', 'else', 'painful', 'weak', 'uneven', 'taut',

			'willing', 'deserving']	'unimportant', 'accurate']	'subdued', 'natured']	'indispensabl e']		'likely', 'hearty']
--	--	--	----------------------------	-------------------------------	--------------------------	----------------------	--	------------------------

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['ed', 'embalmed', 'den', 'tailed', 'weed', 'fell', 'heard', 'ted', 'staked']	['god', 'mark', 'nonsense', 'bible', 'eve', 'curiosity', 'que', 'manor', 'betrayal']	['scrap', 'times', 'trades', 'tickets', 'inks', 'cattle', 'corps', 'brigades', 'moors']	['order', 'department', 'president', 'press', 'director', 'attorney', 'herald', 'city', 'associates']	['bar', 'horse', 'st.', 'dan', 'derby', 'water', 'tombs', 'street', 'dr']	['day', 'joe', 'money', '@', 'daddy', 'show', 'smith', 'j.', 'd.']	['decoration', 'states', 'e.', 'republic', 'war', 'institute', 'may', 'liberty', 'equality']	['pilot', 'bureau', 'harbor', 'c.', 'august', 'avenue', 'ambergris', 'center', 'bridge']
Adjectives	['scared', 'amused', 'embarrassed', 'relaxed', 'mottled', 'excited', 'subdued', 'surprised', 'thrilled', 'worried']	['prone', 'thorough', 'unmixed', 'familiar', 'dignified', 'distinct', 'slight', 'heraldic', 'formless', 'comfortable']	['thousandth', 'outdoors', 'anywheres', 'hundredthousan d', 'indoors', 'together', 'abhors']	['interested', 'unformed', 'qualified', 'non- union', 'willing', 'justified', 'authentic', 'defeatist', 'conscript', 'insupportable']	['steamy', 'muggy', 'fireproof', 'blond', 'lazy', 'coarse', 'yellowish', 'mellow', 'cheeked', 'taut']	['amusing', 'tame', 'sorry', 'cute', 'else', 'jarring', 'hearty', 'heartfelt', 'kindly', 'any- ']	['methodical', , 'possible', 'differ', 'constipated', 'indispensabl e', 'various', 'necessary', 'entire', 'compelling', 'unfair']	['bona', 'desolate', 'damaging', 'unmoving', 'deserving', 'oblong', 'miniature', 'machinelik e', 'overgrown', , 'isinglass']

1950s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['ben', 'boy', 'sue', 'emma', 'grace', 'brother', 'uncle', 'aunt', 'walter']	['word', '@', 'china', 'god', 'world', 'business', 'kingdom', 'man', 'mission']	['damnations', 'ramey', 'climpkins', 'riverbottom', 'lufkin', 'lovelady', 'santone', 'negrochildren', , 'maidie']	['rock', 'creek', 'milk', 'candle', 'earth', 'pasture', 'statue', 'bird', 'bells']	['goods', 'sisters', 'negroes', 'boys', 'marines', 'cousins', 'kings', 'preparations', 'products']	['sing', 'count', 'git', 'pick', 'ai', 'look', 'sleep', 'hope', 'trust']	['berryben', 'ganchion', 'folner', 'aunty', 'malley', 'swimma', 'charity', 'chessy', 'glee']	['ford', 'saturday', 'texas', 'street', 'highway', 'dallas', 'church', 'theatre', 'city']
Adjectives	['sorry', 'shocked', 'surprised', 'ago', 'startled', 'amused', 'puzzled', 'kindly', 'pleasantly', 'uh']	['aware', 'inevitable', 'involved', 'else', 'confused', 'rid', 'conscious',	['balletic', 'rainstreaked', 'lancelike', 'moonhaunted', , 'blackland', 'lusterless', 'napworn',	['enormous', 'ragged', 'motionless', 'moist', 'dense', 'smooth', 'metallic',	['expensive', 'about', 'available', 'respective', 'numerous', 'remaining', 'together',	['spare', 'overcome', 'grab', 'quit', 'repeat', 'establish', 'conceal', 'seem',	['sodden', 'watery', 'resounding', 'leathery', 'unreal', 'soundless', 'decrepit', 'frail',	['settled', 'earlier', 'gone', 'neighboring', 'driven', 'stationed']

		'previous', 'necessary', 'particular']	'unhung', 'lipsticked', 'possiblewart']	'slender', 'twisted', 'faint']	'equipped', 'mainly']	'produce', 'develop']	'unheard', 'ceaseless']	
--	--	--	---	-----------------------------------	--------------------------	--------------------------	----------------------------	--

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['damnations', 'climpkins', 'lufkin', 'lovelady', 'santone', 'negrochildren', 'byejo', 'sometimesp', 'eulaly']	['word', 'charity', 'god', 'satan', 'prayer', 'world', 'savior', 'man', 'miracle']	['ben', 'berryben', 'boy', 'ganchion', 'folner', 'aunty', 'malley', 'swimma', 'sing']	['glee', 'rock', 'milk', 'follie', 'racket', 'orion', 'bulls', 'candle', 'bell']	['@', 'china', 'bill', 'goods', 'meeting', 'negroes', 'house', 'business', 'year']	['chessy', 'ramey', 'riverbottom', 'huffman', 'priss', 'ora', 'chubb', 'selmers', 'huntsville']	['ford', 'saturday', 'texas', 'creek', 'commissary', 'street', 'highway', 'dallas', 'methodist']	['stovall', 'ola', 'absalom', 'tanner', 'jimbob', 'maidie', 'birdie', 'sanderson', 'camilla']
Adjectives	['balletic', 'rainstreaked', 'moonhaunted', 'blackland', 'napworn', 'possiblewart', 'systemsdiscovere d', 'frozenp', 'unshot', 'lemongreen']	['resoundin g', 'unreal', 'aware', 'frail', 'unheard', 'pitiful', 'fluent', 'anxious', 'inevitable', 'futile']	['aghast', 'goin', 'fightin', 'excited', 'scared', 'bewildered', 'tellin', 'tryin', 'sorry', 'tearful']	['sodden', 'watery', 'leathery', 'soundless', 'enormous', 'ragged', 'motionless', 'ceaseless', 'ornamented', 'drowsy']	['involved', 'unchanged', 'decisive', 'overdue', 'unfit', 'previous', 'necessary', 'secure', 'worthwhile', 'dictatorial']	['lacierlike', 'lusterless', 'unhung', 'lipsticked', 'torpid', 'bodied', 'flaccid', 'standoffish', 'crawlin', 'directionles s']	['spacious', 'unmolested', 'hundredth', 'victorious', 'fashionable', 'sheltered', 'transatlantic', 'ago', 'unexplored', 'trackless']	['decrepit', 'brokenhearte d', 'blase', 'acrost', 'killin', 'hurтин', 'paa ahs', 'nekkid', 'speakin', 'bringin']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['damnations', 'ramey', 'climpkins', 'stovall', 'lufkin', 'lovelady', 'santone', 'negrochildren', 'christy']	['ben', 'ford', 'texas', 'milk', 'highway', 'walter', 'theatre', 'florida', 'bell']	['china', 'saturday', 'sing', 'rock', 'grace', 'church', 'brother', 'uncle', 'aunt']	['charity', 'creek', 'pet', 'warren', 'sisters', 'candle', 'queens', 'prayer', 'mill']	['berryben', 'ganchion', 'folner', 'aunty', 'malley', 'swimma', 'chessy', 'glee', 'ola']	['@', 'day', 'p', 'man', 'may', 'us', 'will', 'people', 'time']	['sue', 'emma', 'dallas', 'methodist', 'racket', 'rob', 'bailey', 'pasture', 'memphis']	['boy', 'word', 'street', 'god', 'city', 'bill', 'land', 'house', 'world']
Adjectives	['balletic', 'rainstreaked', 'moonhaunted', 'blackland', 'lusterless', 'napworn', 'unhung', 'unreal', 'possiblewart', 'systemsdiscovere d']	['enormous', , 'anxious', 'smooth', 'awful', 'goin', 'expensive', 'healthy', 'satisfied', 'twisted', 'excited']	['involved', 'frightened', 'particular', 'impossible', 'unable', 'responsible', 'convinced', 'surprised', 'interesting', 'determined']	['inevitable', 'slender', 'prominent', 'endless', 'weary', 'confused', 'rid', 'spoiled', 'jealous', 'civilized']	['sodden', 'watery', 'lacierlike', 'resounding', 'leathery', 'soundless', 'decrepit', 'lipsticked', 'unheard', 'ceaseless']	['about', 'still', 'through']	['ragged', 'motionless', 'frail', 'moist', 'dense', 'futile', 'momentary', 'accompanying', 'rotten', 'careless']	['aware', 'else', 'sorry', 'necessary', 'ago', 'done', 'almost', 'quite', 'around', 'gone']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['sing', 'meeting', 'king', 'ding', 'perishing', 'bowling', 'lightning', 'pal', 'marching']	['boy', 'word', '@', 'brother', 'god', 'uncle', 'sam', 'mama', 'prayer']	['china', 'texas', 'creek', 'aunt', 'theatre', 'postoffice', 'birdie', 'queens', 'bonnie']	['ben', 'glee', 'rock', 'pet', 'milk', 'street', 'highway', 'grace', 'jack']	['seven', 'ned', 'ed', 'gran', 'ted', 'creed', 'seventeen', 'fourteen', 'den']	['ford', 'church', 'warren', 'equinox', 'savior', 'statue', 'business', 'college', 'school']	['goods', 'sisters', 'land', 'negroes', 'mill', 'marines', 'cousins', 'aunts', 'uncles']	['charity', 'commissary', 'sue', 'city', 'bill', 'world', 'year', 'princess', 'preparations']
Adjectives	['resounding', 'stunning', 'terrifying', 'soaring', 'pleasing', 'reassuring', 'unending', 'excruciating', 'distressing', 'amazing']	['sodden', 'soundless', 'motionless', 'aware', 'unheard', 'pitiful', 'anxious', 'timeless', 'feverish', 'mournful']	['lacelike', 'lusterless', 'aghastr', 'moist', 'wiry', 'torpid', 'rubbery', 'flaccid', 'standoffish', 'brittle']	['watery', 'leathery', 'frail', 'drowsy', 'shrunk', 'smooth', 'crawl', 'lightless', 'rotten', 'blase']	['ragged', 'lipsticked', 'ornamented', 'untouched', 'brokenhearted', 'gloved', 'unmolested', 'whiskered', 'talented', 'satisfied']	['unreal', 'enormous', 'decrepit', 'ceaseless', 'spacious', 'treacherous', 'fluent', 'inevitable', 'futile', 'momentary']	['hundredth', 'weighty', 'unaware', 'hundredthousand', 'thees', 'quare', 'neces', 'thousandth', 'millionth', 'nonmembers']	['spectral', 'dense', 'involved', 'expensive', 'prominent', 'bestial', 'intricate', 'solitary', 'decisive', 'continual']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['boy', 'word', 'ford', 'sing', 'grace', 'brother', 'god', 'uncle', 'aunt']	['sue', 'world', 'equinox', 'kingdom', 'hare', 'manor', 'princess', 'miracle', 'next']	['commissary', 'city', 'postoffice', 'bill', 'meeting', 'business', 'college', 'year', 'school']	['goods', 'sisters', 'queens', 'negroes', 'boys', 'marines', 'aunts', 'uncles', 'kings']	['ned', 'ed', 'ted', 'den', 'ken', 'med', 'drew', 'mitten', 'wed']	['@', 'texas', 'highway', 'church', 'theatre', 'hotel', 'racket', 'seven', 'land']	['ben', 'china', 'glee', 'rock', 'creek', 'pet', 'milk', 'street', 'jack']	['charity', 'savior', 'preparations', 'products', 'judgment', 'may', 'peace', 'establishment', 'sis']
Adjectives	['unheard', 'mournful', 'rotten', 'awful', 'hurtin', 'go', 'nasty', 'speakin', 'else', 'hateful']	['resounding', 'lusterless', 'unreal', 'soundless', 'enormous', 'decrepit', 'aware', 'ceaseless', 'spacious', 'treacherous']	['accompanying', 'prominent', 'bestial', 'victorious', 'previous', 'worthwhile', 'unsupervised', 'dictatorial', 'gratified', 'prompt']	['hundredthousand', 'outdoors', 'thees', 'thousandth', 'together', 'upwards', 'bathos', 'popula', 'profes', 'mailmen']	['ragged', 'lipsticked', 'ornamented', 'untouched', 'brokenhearted', 'gloved', 'unmolested', 'whiskered', 'talented', 'barefooted']	['lacelike', 'aghastr', 'torpid', 'standoffish', 'stunning', 'paa-ahs', 'nekkid', 'fightin', 'bringin', 'putrid']	['sodden', 'watery', 'leathery', 'motionless', 'frail', 'drowsy', 'moist', 'wiry', 'shrunk']	['inevitable', 'involved', 'decisive', 'continual', 'maternal', 'inimical', 'necessary', 'contradictory', 'particular', 'secure']

1960s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['millstone', 'ig', 'abrice', 'elisha', 'hubbards', 'madres', 'sierraskeep']	['mount', 'colorado', 'park', 'sunday', 'wednesday', 'los', 'angeles']	['base', 'kitchen', 'creek', 'firing', 'mill', 'plant', 'ground', 'band', 'ducks']	['cope', 'kiss', 'go', 'watch', 'travel', 'show', 'drive', 'look', 'bury']	['tom', 'charlie', 'jesus', 'ben', 'jessup', 'leroy', 'bible', 'perkins', 'sylvester']	['god', 'investment', 'time', 'china', 'press', 'europe', '@', 'war', 'logic']	['opal', 'eunice', 'afterward', 'raunce', 'asa', 'syl', 'levis', 'baldy', 'gibraltar']	['elders', 'indians', 'insurance', 'americans', 'services', 'allies']

	'sylvesterperkins', 'quakeress']	'company', 'christmas']						'germans', 'years', 'ages']
Adjectives	['initiatory', 'big-bosomed', 'marriageand', 'healthful', 'irreligious', 'unpremeditated', 'ante-over', 'sickbed', 'sun-soaked', 'youneed']	['few', 'neighboring']	['homemade', 'comfortable', 'ragged', 'smooth', 'motionless', 'huddled', 'worn', 'muffled', 'taut', 'bulging']	['willing', 'impressed', 'interested', 'worried', 'convinced', 'devoted', 'prepared', 'grateful', 'ashamed', 'satisfied']	['pleased', 'surprised', 'sorry', 'startled', 'delighted', 'puzzled', 'embarrassed', 'ago', 'kidding', 'frankly']	['possible', 'much', 'particular', 'likely', 'reasonable', 'usual', 'profound', 'widespread', 'evident', 'aggressive']	['unmoving', 'emotionless', 'anguished', 'careless', 'miserable', 'forbidding', 'surprising', 'unhappy', 'punctual', 'reliable']	['alone', 'available', 'various', 'numerous', 'respective', 'identical', 'corporate', 'skilled', 'ranging']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['abrice', 'hubbards', 'madres', 'sierraskeep', 'sylvesterperkins', 'quakeress', 'wendlin'sjaunty', 'raunces', 'suetenecan']	['millstone', 'levis', 'reservoir', 'scars', 'glint', 'kitchen', 'firing', 'ground', 'carrot']	['father', 'wilson', 'louella', 'wagner', 'julian', 'nov', 'lewis', 'e', 'franklin']	['elders', 'bible', 'quakers', 'logic', 'mormons', 'rapture', 'blush', 'revelations', 'sin']	['tom', 'opal', 'eunice', 'charlie', 'afterward', 'jesus', 'raunce', 'ben', 'jessup']	['investment', 'company', 'insurance', 'time', 'cope', 'press', '@', 'war', 'plant']	['ig', 'elisha', 'baldy', 'jessups', 'irby', 'ortiz', 'cudlips', 'fitzgeralds', 'moun']	['mount', 'colorado', 'park', 'sunday', 'wednesday', 'indians', 'los', 'angeles', 'christmas']
Adjectives	['initiatory', 'sun-soaked', 'youneed', 'womanless', 'soapweed', 'welcomeand', 'untired', 'alecky', 'mentionable', 'partyless']	['unmoving', 'homemade', 'ripe', 'silvery', 'graying', 'skinned', 'fleshy', 'thirsty', 'grimy', 'frilly']	['stocky', 'bespectacled', 'terse', 'indefatigable', 'uncontested', 'biweekly', 'congratulatory', 'outgoing', 'taciturn', 'bearded']	['anguished', 'careless', 'miserable', 'alone', 'forbidding', 'surprising', 'unhappy', 'punctual', 'alike', 'penetrating']	['aback', 'pleased', 'amazed', 'surprised', 'excited', 'shocked', 'sorry', 'pretty', 'afterthought', 'tired']	['willing', 'possible', 'reliable', 'economical', 'available', 'suitable', 'wholehearted', 'likely', 'reasonable', 'pertinent']	['emotionless', 'big-bosomed', 'marriageand', 'unusable', 'cannibal', 'irreligious', 'unpremeditated', 'ante-over', 'sickbed', 'discourteous']	['few', 'somewheres', 'cross-country', 'sunset', 'balmy', 'mountainous', 'itinerant', 'secluded', 'moribund', 'tumultuous']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['opal', 'eunice', 'abrice', 'jessup', 'asa', 'hubbards', 'sierra', 'madres', 'sierraskeep']	['tom', 'charlie', 'jesus', 'park', 'bible', 'indians', 'investment', 'wagner', 'los', 'angeles']	['mount', 'afterward', 'colorado', 'wednesday', 'wagner', 'pete', 'nov', 'lewis', 'ellen']	['god', 'father', 'war', 'house', 'world', 'york', 'place', 'years', 'mother']	['millstone', 'ig', 'raunce', 'elisha', 'syl', 'sylvester', 'levis', 'baldy', 'gibraltar']	['ben', 'sunday', 'company', 'california', 'mary', 'china', 'e', 'lieutenant', 'press']	['leroy', 'elders', 'perkins', 'lowe', 'ramos', 'julian', 'franklin', 'hannah', 'luther']	['time', '@', 'go', 'may', 'one', 'will', 'people', 'man', 'd']

Adjectives	['initiatory', 'anguished', 'careless', 'big- bosomed', 'marriageand', 'penetrating', 'homemade', 'unmarried', 'eccentric', 'unreal']	['interesting', 'pleased', 'reasonable', 'worried', 'convinced', 'attractive', 'comfortable', 'evident', 'startled', 'delighted']	['miserable', 'surprising', 'unhappy', 'reliable', 'alike', 'ignorant', 'impressed', 'suitable', 'excited', 'shocked']	['alone', 'possible', 'few', 'along', 'ago', 'suddenly', 'rather', 'ever', 'between', 'trying']	['unmoving', 'emotionless', 'forbidding', 'punctual', 'healthful', 'aback', 'unusable', 'cannibal', 'wholehearted', 'unpremeditate d']	['willing', 'available', 'surprised', 'particular', 'likely', 'interested', 'sorry', 'pretty', 'tired', 'usual']	['economical', 'ripe', 'amazed', 'helpful', 'uneasy', 'fearful', 'embarrassing', 'profound', 'disgusted', 'uncomfortable']	['much', 'about', 'going', 'too', 'before']
-------------------	--	--	---	---	--	---	--	--

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['millstone', 'afterward', 'ig', 'e', 'christ', 'lieutenant', 'blackbird', 'j', 'chairman']	['company', 'levis', 'reservoir', 'cope', 'ana', 'depot', 'church', 'logic', 'fort']	['god', 'bible', 'los', 'father', 'time', 'fry', '@', 'fe', 'cafe']	['elders', 'perk', 'scars', 'mill', 'plant', 'aids', 'corps', 'marines', 'island']	['tom', 'mount', 'ben', 'park', 'valley', 'bob', 'camp', 'china', 'rosa']	['firing', 'und', 'landing', 'king', 'building', 'end', 'shipping', 'barring', 'ding']	['ven', 'greed', 'ken', 'ted', 'wed', 'spotted', 'ten', 'bled', 'abed']	['opal', 'investment', 'insurance', 'press', 'base', 'war', 'inc', 'al', 'ind']
Adjectives	['homemade', 'ante-over', 'outspoken', 'benign', 'picturesque', 'collarless', 'grimy', 'masculine', 'forthright', 'spicy']	['miserable', 'punctual', 'healthful', 'ignorant', 'unusable', 'eccentric', 'unreal', 'dissatisfied', 'suitable', 'cannibal']	['emotionless', 'careless', 'alone', 'unhappy', 'alike', 'much', 'helpful', 'uneasy', 'fearful', 'sorry']	['somewhere s', 'mild', 'groundless', 'thousandth', 'indoors', 'millionth', 'neces', 'between', 'unawares', 'one-acters']	['aback', 'ripe', 'silvery', 'fleshy', 'cloudless', 'watery', 'scraggly', 'handmade', 'shadowy', 'smooth']	['unmoving', 'willing', 'forbidding', 'surprising', 'penetrating', 'pleasing', 'graying', 'embarrassin g', 'easygoing', 'lingering']	['anguished', 'few', 'unmarried', 'impressed', 'pleased', 'amazed', 'surprised', 'excited', 'wholehearted', 'shocked']	['possible', 'reliable', 'interesting', 'economical', 'available', 'particular', 'attractive', 'indecent', 'intimate', 'evident']

s

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['millstone', 'ig', 'reservoir', 'camp', 'e', 'papa', 'lieutenant', 'blackbird', 'horseback']	['opal', 'investment', 'company', 'insurance', 'ana', 'press', 'chairman', 'railway', 'church']	['tom', 'mount', 'ben', 'park', 'valley', 'bob', 'china', 'rosa', 'shel']	['afterward', 'god', 'bible', 'los', 'father', 'time', 'perk', 'christ', 'fe']	['elders', 'scars', 'aids', 'corps', 'marines', 'village', 'wheat', 'queens', 'coast']	['levis', 'cope', 'logic', 'inc', 'al', 'ind', 'sis', 'revelations', 'world']	['greed', 'ken', 'ted', 'wed', 'spotted', 'bled', 'abed', 'kitten', 'struck']	['firing', 'landing', 'king', 'building', 'shipping', 'barring', 'ding', 'dong', 'riding']
Adjectives	['few', 'homemade', 'cannibal', 'ante- over', 'homesick', 'easygoing', 'picturesque',	['outspoken', 'devoted', 'prepared', 'unwarranted', 'undemonstrativ e', 'discouraged', 'exemplary',	['ripe', 'silvery', 'fleshy', 'collarless', 'grimy', 'cloudless', 'spicy',	['emotionless , 'careless', 'miserable', 'alone', 'surprising', 'unhappy', 'alike',	['catty', 'scanty', 'thousandth', 'indoors', 'carbolic', 'carbonic', 'hig', 'larly', 'neces', 'rests']	['possible', 'punctual', 'reliable', 'interesting', 'economical', 'available', 'unusable',	['anguished', 'unmarried', 'impressed', 'pleased', 'amazed', 'surprised', 'excited',	['unmoving', 'willing', 'forbidding', 'penetrating', 'pleasing', 'graying', 'lingering',

	'cross-country', 'lackluster', 'neighborly']	'illicit', 'despondent', 'involuntary']	'watery', 'tangy', 'colorless']	'much', 'healthful', 'ignorant']		'dissatisfied', 'suitable', 'particular']	'shocked', 'skinned', 'troubled']	'misleading', 'arousing', 'appealing']
--	--	---	---------------------------------------	--	--	---	---	--

1970s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['otto', 'sophie', 'charlie', 'ruth', 'paul', 'bush', 'henry', 'garland', 'judy']	['holsteins', 'meissen', 'iglesia', 'pucci', 'nhat', '12a', 'rohe', 'soph', 'youngmike']	['god', '@', 'world', 'family', 'court', 'pain', 'renaissance', 'disorder', 'guide']	['brooklyn', 'heights', 'street', 'broadway', 'june', 'city', 'avenue', 'island', 'september']	['health', 'will', 'state', 'vehicle', 'work', 'commerce', 'may', 'commission', 'budget']	['wall', 'fire', 'gin', 'darkness', 'pan', 'earth', 'sea', 'cup', 'table']	['catholics', 'americans', 'jews', 'negroes', 'pages', 'witnesses', 'mothers', 'greeks', 'veterans']	['bentwood', 'tiffany', 'goethe', 'flynders', 'muni', 'kleenex', 'russel', 'bowery', 'myron']
Adjectives	['surprised', 'pleased', 'dear', 'amused', 'startled', 'shocked', 'impressed', 'annoyed', 'kidding', 'oh']	['jowled', 'inp5the', 'mumbleof', 'gettingcold', 'ani- mal', 'exterin', 'mercedesbenzsed an', 'soporific', 'endbrought', 'reductive']	['about', 'else', 'conscious', 'afraid', 'proud', 'particular', 'intense', 'profound', 'vague', 'subtle']	['alone', 'later', 'driven', 'assembled']	['likely', 'involved', 'such', 'further', 'substantial', 'determined', 'important', 'reasonable', 'unable', 'available']	['transparent', 'slight', 'coarse', 'neat', 'plump', 'downward', 'stuck', 'damp', 'shiny', 'wheeled']	['various', 'alike', 'numerous', 'remaining', 'together', 'mostly']	['fluorescent', 'scruffy', 'unprincipled', 'starved', 'anguished', 'stunned', 'horried', 'astonished', 'despicable', 'livid']

Word2vec (Skip- gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['iglesia', 'nhat', '12a', 'soph', 'youngmike', 'pluvieux', 'someonehad', 'towardwhich', 'jamdyce']	['catholics', 'matisse', 'jews', 'world', 'family', 'rabies', 'pain', 'renaissance', 'hitler']	['otto', 'sophie', 'god', 'charlie', 'ruth', 'garland', 'judy', 'myron', 'flo']	['holsteins', 'meissen', 'pucci', 'muni', 'holstein', 'mies', 'rohe', 'cervantes', 'baudelaire']	['@', 'americans', 'institute', 'court', 'witnesses', 'health', 'citizen', 'veterans', 'carter']	['bentwood', 'tiffany', 'wall', 'kleenex', 'gin', 'styrofoam', 'paisley', 'pall', 'buick']	['goethe', 'june', 'paul', 'henry', 'russel', 'van', 'anne', 'gottlieb', 'luke']	['brooklyn', 'heights', 'street', 'broadway', 'flynders', 'city', 'bush', 'avenue', 'island']
Adjectives	['jowled', 'inp5the', 'mumbleof', 'gettingcold', 'ani- mal', 'exterin', 'mercedesbenzsed an', 'endbrought', 'preferencesand', 'orientalgenetic']	['starved', 'conscious', 'despicable', 'unreasonable', 'superficial', 'tiresome', 'comfortable', 'racist', 'sympathetic', 'uncharacteristic']	['surprised', 'else', 'anguished', 'stunned', 'horried', 'astonished', 'puzzled', 'sidelong', 'nauseated', 'awful']	['scruffy', 'unprinciple d', 'unlovely', 'soporific', 'reductive', 'pre- columbian', 'hieratic', 'vertiginous', ,	['about', 'likely', 'involved', 'such', 'unlawful', 'further', 'substantial', 'important', 'unprecedente d', 'reasonable']	['fluorescent', 'transparent', 'livid', 'chalky', 'pudgy', 'unmoving', 'murky', 'slight', 'limber', 'coarse']	['expatriate', 'uncontested', 'personable', 'unsuccessful', , 'later', 'enterprising', 'unimpressed', , 'portly', 'affable', 'hapless']	['unmarked', 'populous', 'desolate', 'prosperous', 'unoccupied', 'squalid', 'unmanned', 'sprawling', 'secluded', 'alone']

				'conciliatin g', 'snotty']				
--	--	--	--	-------------------------------	--	--	--	--

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['sophie', 'heights', 'catherine', 'suzanne', 'emily', 'negroes', 'renaissance', 'oakland', 'sixty']	['meissen', 'iglesia', 'pucci', 'nhat', '12a', 'garland', 'rohe', 'soph', 'youngmike']	['god', 'city', 'world', 'family', 'house', 'york', 'day', 'word', 'children']	['wall', 'street', 'henry', 'americans', 'island', 'london', 'march', 'friday', 'fire']	['brooklyn', 'june', 'charlie', 'ruth', 'paul', 'avenue', 'september', 'queen', 'mike']	['bentwood', 'tiffany', 'goethe', 'flynders', 'holsteins', 'muni', 'kleenex', 'russel', 'bowery']	['@', 'will', 'time', 'man', 'i', 'a', 'to', 'an', 'can']	['otto', 'broadway', 'catholics', 'bush', 'judy', 'van', 'anne', 'bernard', 'fred']
Adjectives	['transparent', 'stunned', 'horrificed', 'unreasonable', 'sympathetic', 'frantic', 'insulting', 'potent', 'irritated', 'coarse']	['jowled', 'inp5the', 'starved', 'mumbleof', 'gettingcold', 'ani- mal', 'superficial', 'racist', 'exterin', 'mercedesbenzsed an']	['else', 'such', 'afraid', 'alone', 'later', 'trying', 'done', 'ever', 'around', 'together']	['surprised', 'likely', 'involved', 'tired', 'further', 'worried', 'important', 'particular', 'available', 'necessary']	['conscious', 'comfortable', 'awful', 'slight', 'determined', 'scared', 'proud', 'frightened', 'excited', 'unable']	['fluorescent', 'scruffy', 'unprincipled', 'anguished', 'despicable', 'livid', 'tiresome', 'uncharacterist ic', 'unlovely', 'unloved']	['about', 'on']	['astonished', 'puzzled', 'helpless', 'substantial', 'threatening', 'accustomed', 'peculiar', 'reasonable', 'inclined', 'occasional']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['city', 'institute', 'court', 'health', 'canal', 'depression', 'religion', 'humanism', 'age']	['bentwood', 'wall', 'street', 'bush', 'st.', 'bay', 'fire', 'river', 'pain']	['king', 'reading', 'building', 'wandering', 'ring', 'ding', 'tumbling', 'working', 'engineering']	['ruth', 'avenue', 'garland', 'van', 'panama', 'army', 'squire', 'bobby', 'locust']	['god', '@', 'queen', 'mike', 'sir', 'lord', 'madame', 'mein', 'jell']	['ned', 'ted', 'ed', 'ken', 'illustrated', 'raid', 'reed', 'den', 'published']	['heights', 'island', 'rabies', 'fairies', 'pages', 'witnesses', 'mall', 'veterans', 'slaves']	['world', 'vieux', 'march', 'family', 'renaissance', 'theater', 'library', 'althoughshe', 'concord']
Adjectives	['superficial', 'involved', 'such', 'furtive', 'reductive', 'hieratic', 'substantial', 'important', 'banal', 'unimportant']	['sidelong', 'slight', 'limber', 'coarse', 'stricken', 'shadowy', 'neat', 'shabby', 'sallow', 'frail']	['insulting', 'unmoving', 'appalling', 'threatening', 'extinguishing', , 'irritating', 'pleasing', 'unappetizing', 'disfiguring', 'appealing']	['fluorescen t', 'scruffy', 'tiresome', 'chalky', 'pudgy', 'murky', 'ladylike', 'buoyant', 'stubborn', 'snotty']	['about', 'else', 'livid', 'comfortable', 'unlovely', 'awful', 'pitiful', 'helpless', 'childlike', 'pointless']	['surprised', 'starved', 'anguished', 'stunned', 'horrificed', 'astonished', 'puzzled', 'unloved', 'nauseated', 'unmarked']	['groundless', 'thousandth', 'neces', 'ques', 'unawares', 'countable', 'indoors', 'non- members', 'rests', 'scares']	['unprincipled', , 'transparent', 'conscious', 'despicable', 'likely', 'unreasonable', 'racist', 'sympathetic', 'uncharacterist ic', 'frantic']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['world', 'renaissance', 'althoughshe', 'concord', 'inn', 'guide', 'titanic', 'depression', 'destiny']	['bentwood', 'wall', 'street', 'bush', 'van', 'mike', 'st.', 'bay', 'fire']	['institute', 'health', 'disorder', 'religion', 'humanism', 'justice', 'equality', 'art', 'revolution']	['ned', 'ted', 'ed', 'ken', 'reed', 'drew', 'deflated', 'arrested', 'brokenly']	['heights', 'rabies', 'fairies', 'pages', 'mothers', 'slaves', 'carnations', 'children', 'hogs']	['god', 'ruth', 'queen', 'pain', 'shaker', 'munch', 'sir', 'roach', 'chateau']	['city', 'family', 'court', 'army', 'witnesses', 'citizen', 'veterans', 'will', 'state']	['@', 'avenue', 'island', 'garland', 'vieux', 'march', 'panama', 'bobby', 'locust']
Adjectives	['transparent', 'conscious', 'despicable', 'livid', 'tiresome', 'comfortable', 'sympathetic', 'such', 'furtive', 'frantic']	['fluorescent', 'scruffy', 'chalky', 'unmoving', 'murky', 'slight', 'limber', 'coarse', 'snotty', 'sunburnt']	['unprincipled', 'unreasonable', 'superficial', 'uncharacteristic', 'reductive', 'hieratic', 'further', 'substantial', 'important', 'habitual']	['surprised', 'starved', 'anguished', 'stunned', 'horrificed', 'astonished', 'puzzled', 'unloved', 'nauseated', 'unmarked']	['thousandth', 'hundredth', 'outdoorsmen', 'together', 'scattershot', 'mostly', 'indoors', 'anywheres', 'ranging', 'rests']	['about', 'else', 'unlovely', 'awful', 'pudgy', 'helpless', 'childlike', 'ladylike', 'afraid', 'stubborn']	['likely', 'racist', 'involved', 'plaintive', 'secretive', 'banal', 'unprecedented', 'available', 'ignominious', 'convincing']	['sidelong', 'buoyant', 'spongy', 'flutulent', 'sketchy', 'cunning', 'hourglass', 'misshapen', 'stocky', 'acrid']

1980s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['dobbin', 'thackeray', 'sneden', 'shc', 'boop', 'pickf', 'Tetranger', 'p112parker', 'todayit']	['pick', 'come', 'kiss', 'pass', 'cope', 'drop', 'cook', 'cry', 'play']	['mary', 'paul', 'nick', 'cynthia', 'nina', 'lord', 'taylor', 'ellis', 'gene']	['york', 'landing', 'friday', 'korea', 'road', 'holiday', 'orchestra', 'washington', 'delaware']	['metcalf', 'billie', 'alto', 'quill', 'scouts', 'odyssey', 'landers', 'spangle', 'madrid']	['vanity', '@', 'stranger', 'world', 'war', 'god', 'america', 'times', 'bible']	['chocolate', 'coke', 'chicken', 'statue', 'cadillac', 'sun', 'animal', 'jaws', 'helmet']	['states', 'adults', 'bond', 'may', 'americans', 'europe', 'christians', 'will', 'chiefs']
Adjectives	['almostbroke', 'monkeyswere', 'drivingfast', 'prophylactic', 'ten-speed', 'sedentary', 'illustrateswhy', 'insideand', 'dumpy', 'unsociable']	['refuse', 'distinguish', 'repeat', 'abandon', 'overcome', 'pretend', 'enter', 'swear', 'resolve', 'enjoy']	['surprised', 'embarrassed', 'pleased', 'startled', 'kindly', 'amused', 'puzzled', 'kidding', 'ago', 'versa']	['outside', 'crowded', 'outdoor', 'parked', 'farther', 'deserted', 'wooded', 'driven', 'adjoining', 'neighboring']	['eager', 'tired', 'ludicrous', 'perceptive', 'anxious', 'irrelevant', 'condescending', 'predictable', 'ironic', 'comparable']	['else', 'much', 'painful', 'afraid', 'obvious', 'aware', 'interested', 'about', 'logical', 'capable']	['shiny', 'outstretched', 'enormous', 'sturdy', 'massive', 'delicate', 'gigantic', 'downward', 'moist', 'horizontal']	['possible', 'willing', 'able', 'involved', 'adequate', 'various', 'additional', 'necessary', 'substantial', 'unable']

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
----------------------	----	----	----	----	----	----	----	----

Names	['dobbin', 'sneden', 'boop', 'pickf', 'tetranger', 'p112parker', 'todayit', 'putrefy', 'salems']	['vanity', 'freud', 'adults', 'world', 'cleopatra', 'afterlife', 'liberty', 'antics', 'miracles']	['paul', 'cynthia', 'taylor', 'billie', 'ellis', 'orchestra', 'alto', 'gene', 'odyssey']	['mary', 'metcalf', 'nick', 'nina', 'lord', 'boy', 'landers', 'louise', 'madrid', 'spangle']	['york', 'landing', 'friday', 'road', 'holiday', 'scouts', 'delaware', 'madrid', 'haven']	['korea', 'washington', 'states', '@', 'war', 'philippines', 'republic', 'bond', 'america']	['thackeray', 'shc', 'ord', 'camus', 'uh-uh', 'nab', 'garcon', 'darien', 'antony']	['quill', 'chocolate', 'coke', 'styrofoam', 'chicken', 'perrier', 'rye', 'sun', 'hulk']
Adjectives	['almostbroke', 'monkeyswere', 'drivingfast', 'illustrateswhy', 'too-', 'opposedto', 'goddamn', 'magician', 'eccentric', 'cold', 'courtroom', 'the', 'somethingalong']	['eager', 'ludicrous', 'perceptive', 'much', 'irrelevant', 'painful', 'condescending', 'predictable', 'ironic', 'harmless']	['born-again', 'itinerant', 'miffed', 'lanky', 'thirtieth', 'prominent', 'portly', 'hawkish', 'estranged', 'outgoing']	['tired', 'else', 'anxious', 'surprised', 'embarrassed', 'afraid', 'spooky', 'frightened', 'scared', 'awful']	['consecutive', 'impassable', 'outside', 'unarmed', 'ragtag', 'massed', 'crowded', 'breakneck', 'spacious', 'treeless']	['comparable', 'possible', 'willing', 'able', 'involved', 'adequate', 'about', 'reasonable', 'unequal', 'similar']	['certifiable', 'prophylactic', 'ten-speed', 'insideand', 'dumpy', 'unsociable', 'blooded', 'non-jew', 'soporific', 'gawky']	['shiny', 'pubic', 'sweaty', 'ripe', 'sunken', 'heeled', 'overstuffed', 'fluorescent', 'lumpy', 'bloodshot']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['dobbin', 'sneden', 'cynthia', 'ellis', 'scouts', 'odyssey', 'madrid', 'boop', 'pickf']	['friday', 'ann', 'california', 'christ', 'freud', 'pick', 'joel', 'saturday', 'paris']	['mary', 'york', 'paul', 'lord', 'road', 'boy', 'washington', 'states', 'john']	['vanity', 'nina', 'orchestra', 'delaware', 'betty', 'madison', 'whitney', 'chocolate', 'easter']	['@', 'one', 'a', 'i', 'for', 'have', 'all', 's', 'would']	['night', 'world', 'come', 'god', 'times', 'p', 'may', 'father', 'will']	['landing', 'nick', 'korea', 'taylor', 'holiday', 'gene', 'louise', 'haven', 'yale']	['thackeray', 'metcalf', 'billie', 'alto', 'quill', 'landers', 'shc', 'spangle', 'houdini']
Adjectives	['almostbroke', 'monkeyswere', 'lousy', 'sweaty', 'drivingfast', 'aback', 'mindless', 'sunken', 'depressing', 'consecutive']	['eager', 'willing', 'obvious', 'frightened', 'scared', 'healthy', 'capable', 'confused', 'similar', 'various']	['tired', 'surprised', 'afraid', 'possible', 'able', 'aware', 'interested', 'involved', 'worried', 'particular']	['irrelevant', 'predictable', 'ironic', 'harmless', 'embarrassing', 'shiny', 'depressed', 'ripe', 'disgusted', 'mistaken']	['about']	['else', 'much', 'done', 'away', 'together', 'ever', 'used', 'around', 'known', 'ago']	['anxious', 'painful', 'embarrassed', 'comparable', 'awkward', 'awful', 'uncomfortable', 'amazed', 'adequate', 'disturbed']	['ludicrous', 'perceptive', 'condescending', 'spooky', 'incorrect', 'certifiable', 'pubic', 'creepy', 'quirky', 'insensitive']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['landing', 'ring', 'flying', 'king', 'sleeping', 'spring', 'riding', 'thanksgiving', 'evening']	['vanity', 'states', 'sal', 'disco', 'world', 'war', 'city', 'march', 'republic']	['scouts', 'landers', 'adults', 'heights', 'miracles', 'village', 'times']	['san', 'ed', 'ted', 'seed', 'wed', 'hed', 'creed', 'bed', 'childbed']	['alto', 'ord', 'superman', 'kingdom', 'statue', 'liberty', 'circus', 'animal', 'theater']	['nick', 'road', 'pick', 'street', 'night', 'coke', 'park', 'heaven', 'farm']	['holiday', 'orchestra', 'odyssey', 'ann', 'john', 'avenue', 'nab', 'chocolate', 'august']	['lord', 'gene', 'quill', 'boy', '@', 'haven', 'stranger', 'come', 'god']

			'tales', 'county']					
Adjectives	['condescending', 'willing', 'embarrassing', 'exciting', 'depressing', 'caring', 'overpowering', 'overwhelming', 'liberating', 'distressing']	['perceptive', 'comparable', 'incorrect', 'interested', 'involved', 'adequate', 'insensitive', 'consecutive', 'sadistic', 'prophylactic']	['millionth', 'thousandth', 'banty', 'oilmen', 'unawares', 'neces', 'underclothes ']	['tired', 'surprised', 'embarrassed', , 'frightened', 'scared', 'amazed', 'depressed', 'disturbed', 'ashamed', 'worried']	['ludicrous', 'irrelevant', 'predictable', 'ironic', 'possible', 'obvious', 'certifiable', 'pubic', 'unbelievable', 'expressionless ']	['shiny', 'creepy', 'sweaty', 'aback', 'ripe', 'sunken', 'bloodshot', 'shrunk', 'colorful', 'wraparound ']	['eager', 'spooky', 'awkward', 'quirky', 'microscopic', 'ten-speed', 'fluorescent', 'lumpy', 'dumpy', 'puffy']	['else', 'anxious', 'much', 'painful', 'afraid', 'harmless', 'able', 'awful', 'aware', 'uncomfortable ']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['orchestra', 'odyssey', '@', 'pick', 'avenue', 'august', 'march', 'park', 'styrofoam']	['nick', 'lord', 'quill', 'boy', 'haven', 'john', 'ord', 'stranger', 'night']	['scouts', 'landers', 'adults', 'heights', 'boys', 'times', 'tales', 'sentinels', 'lands']	['landing', 'flying', 'king', 'ed', 'sleeping', 'riding', 'evening', 'ted', 'wadding']	['road', 'street', 'chocolate', 'coke', 'chicken', 'sun', 'peter', 'hulk', 'jaws']	['holiday', 'ann', 'states', 'city', 'superman', 'afterlife', 'bond', 'bill', 'college']	['vanity', 'alto', 'gene', 'sal', 'disco', 'world', 'kingdom', 'animal', 'inc']	['war', 'liberty', 'republic', 'antics', 'may', 'testament', 'society', 'acts', 'revolution']
Adjectives	['spooky', 'pubic', 'ten-speed', 'dumpy', 'facile', 'somber', 'gawky', 'outside', 'ragtag', 'die-hard']	['eager', 'tired', 'else', 'anxious', 'surprised', 'embarrassed', 'afraid', 'frightened', 'scared', 'awkward']	['thousandth', 'peopled', 'neces', 'nese']	['amazed', 'worried', 'heeled', 'disgusted', 'overstuffed', 'exhausted', 'stoned', 'caring', 'inclined', 'outstretched ']	['shiny', 'creepy', 'sweaty', 'ripe', 'sunken', 'fluorescent', 'lumpy', 'bloodshot', 'shrunk', 'wraparound']	['willing', 'consecutive', , 'sedentary', 'earlier', 'confident', 'qualified', 'reticent', 'advantaged', 'meager', 'indebted']	['ludicrous', 'much', 'irrelevant', 'painful', 'condescending', 'qualified', 'possible', 'harmless', 'obvious', 'embarrassing']	['perceptive', 'predictable', 'comparable', 'incorrect', 'involved', 'adequate', 'logical', 'reasonable', 'meaningful', 'unequal']

1990s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['valentines', 'pinks', 'rimsky', 'alexandrovich', , 'milovan', 'petroshki', 'ferraris', 'uhggm', 'ohhhhhhhhhhh ']	['rachael', 'melanoma', 'turk', 'oprah', 'dini', 'frisbee', 'jemma', 'hangover', 'pow']	['mission', 'god', '@', 'life', 'death', 'act', 'theory', 'way', 'sin']	['bulbs', 'sun', 'tea', 'toe', 'perfume', 'waters', 'tree', 'dough', 'head']	['rent', 'look', 'fade', 'burn', 'hire', 'wait', 'hell', 'start', 'include']	['day', 'february', 'april', 'hill', 'creek', 'ave', 'india', 'wedding', 'garden']	['rachel', 'joe', 'amy', 'henry', 'cary', 'grant', 'lady', 'ivan', 'prince']	['dragons', 'workers', 'dogs', 'guards', 'brothers', 'words', 'writers', 'sisters', 'fathers']

Adjectives	['oozy', 'unable', 'lumpish', 'constructive', 'decent', 'lusterless', 'omnious', "a'pose", 'criadas', 'pusseywhipped']	['contemplative', 'tentative', 'rapt', 'uncomfortable', '', 'cranky', 'ornery', 'brassy', 'unlocked', 'bitchy', 'feathery']	['much', 'obvious', 'about', 'underlying', 'apparent', 'interested', 'willing', 'profound', 'capable', 'accurate']	['ornate', 'wrinkled', 'thick', 'colorful', 'motionless', 'sweaty', 'swollen', 'salty', 'sleek', 'plump']	['afraid', 'overcome', 'pretend', 'tear', 'giving', 'handed', 'steal', 'boost', 'ruin', 'prevent']	['adjacent', 'earlier', 'sometime', 'sprawling', 'adjoining', 'wooded', 'neighboring', 'leading']	['pleased', 'married', 'amused', 'surprised', 'embarrassed', 'annoyed', 'renowned', 'respected', 'oh', 'yes']	['numerous', 'diverse', 'newly', 'were']
-------------------	---	--	---	--	--	--	---	---

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['pinsk', 'alexandrovich', 'milovan', 'uhggm', 'ohhhhhhhhhh', '', 'ueghh', 'ahhhhyes', 'ohhhhhhhh', 'vereniki']	['rachael', 'melanoma', 'valentines', 'minsk', 'rimsky', 'vasily', 'elenia', 'petroshki', 'ferraris']	['god', 'hangover', 'occident', 'life', 'death', 'fanfare', 'kingdom', 'theory', 'way']	['rachel', 'amy', 'oprah', 'dini', 'jemma', 'pow', 'lady', 'ivan', 'prince']	['bulbs', 'frisbee', 'sun', 'tea', 'toe', 'perfume', 'tree', 'dough', 'spider']	['mission', 'rent', '@', 'workers', 'act', 'bureau', 'rehab', 'circuit', 'year']	['day', 'turk', 'february', 'april', 'hill', 'creek', 'ave', 'dragons', 'zhar']	['joe', 'henry', 'cary', 'grant', 'victor', 'don', 'bueno', 'mara', 'enrique']
Adjectives	['omnious', "a'pose", 'criadas', 'ashame', 'demasiada', 'ndale', 'anti-gravity', 'imperceivable', 'non-skilled', 'swarmy']	['oozy', 'bitchy', 'wolfish', 'lumpish', 'lusterless', 'pusseywhipped', '', 'purty', 'chested', 'hearted', 'legless']	['much', 'contemplative', 'tentative', 'uncomfortable', '', 'obvious', 'endless', 'about', 'childish', 'threatening', 'uneasy']	['rapt', 'cranky', 'ornery', 'unlocked', 'pleased', 'afraid', 'terrified', 'puzzled', 'afterthought', '', 'miserable']	['brassy', 'feathery', 'shadowy', 'ornate', 'foreground', 'rotund', 'wrinkled', 'leathery', 'companionable', '', 'gauzy']	['unemployed', 'involved', 'accountable', 'additional', 'questionable', 'outpatient', 'likely', 'modest', 'staggering', 'unprecedented', '']	['pristine', 'trendy', 'stately', 'treacherous', 'unarmed', 'sleazy', 'rustic', 'tortuous', 'impassable', 'clapboard']	['married', 'bespectacled', '', 'affable', 'talented', 'unsigned', 'fortyish', 'posthumous', '', 'esteemed', 'renowned', 'valedictory']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['pinsk', 'kiev', 'alexandrovich', 'milovan', 'hmm', 'petroshki', 'shhh', 'baba', 'uhggm']	['rachael', 'joe', 'february', 'grant', 'creek', 'prince', 'uncle', 'tea', 'india']	['rent', 'amy', 'toe', 'victor', 'perfume', 'waters', 'devil', 'chile', 'el']	['@', 'way', 'time', 'will', 'man', 'people', "'d", 'one', 'p']	['rachael', 'melanoma', 'valentines', 'turk', 'oprah', 'dini', 'frisbee', 'jemma', 'cary']	['day', 'god', 'life', 'death', 'head', 'look', 'boy', 'year', 'may']	['mission', 'april', 'henry', 'hill', 'sun', 'lady', 'tree', 'garden', 'saturday']	['bulbs', 'dragons', 'ivan', 'teasing', 'koschei', 'bride', 'dough', 'spider', 'heavens']
Adjectives	['oozy', 'shadowy', 'childish', 'incapable', 'depressing', 'lifeless', 'worthless', 'omnious', "a'pose", 'criadas']	['pleased', 'unable', 'decent', 'awake', 'ashamed', 'frightened', 'satisfied', 'apparent', 'impressed', 'capable']	['uncomfortable', 'endless', 'threatening', 'terrified', 'surprising', 'distracted', 'profound', 'peculiar', 'swollen', 'meaningful']	['much', 'about', 'before', 'were']	['contemplative', '', 'rapt', 'cranky', 'ornery', 'brassy', 'bitchy', 'feathery', 'ornate', 'rotund', 'leathery']	['married', 'likely', 'almost', 'around', 'done', 'along', 'behind', 'oh', 'yes', 'ago']	['obvious', 'afraid', 'thick', 'interested', 'worried', 'scared', 'surprised', 'involved', 'particular']	['tentative', 'unlocked', 'foreground', 'wrinkled', 'uneasy', 'arrogant', 'fearful', 'puzzled', 'fascinating', 'miserable']

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['melanoma', 'joe', 'pause', 'frisbee', 'hmm', 'auntie', 'perfume', 'matouchka', 'kingdom']	['grant', 'victor', 'fanfare', 'presto', 'devil', 'wart', 'vas', 'por', 'diablo']	['teasing', 'wedding', 'se', 'humming', 'ding', 'writing', 'dong', 'stealing', 'bang']	['barbed', 'enlisted', 'ed', 'stalled', 'ted', 'teen', 'bed', 'crowed', 'called']	['day', 'god', 'ave', '@', 'hangover', 'lady', 'uncle', 'life', 'death']	['dragons', 'waters', 'groans', 'es', 'workers', 'sabs', 'pues', 'obits', 'bites']	['mission', 'rent', 'prince', 'orient', 'act', 'al', 'est', 'theory', 'prize']	['bulbs', 'hill', 'creek', 'pow', 'sun', 'shhh', 'tea', 'baba', 'toe']
Adjectives	['rapt', 'oozy', 'ornery', 'wolfish', 'ticklish', 'arrogant', 'gauzy', 'lumpish', 'crunchy', 'gaudy']	['obvious', 'ornate', 'endless', 'incapable', 'unable', 'miserable', 'worthless', 'surreal', 'sarcastic', 'impassive']	['threatening', 'depressing', 'overcoming', 'fascinating', 'troubling', 'surprising', 'comforting', 'underlying', 'appalling', 'willing']	['unlocked', 'pleased', 'wrinkled', 'terrified', 'puzzled', 'dejected', 'unconvinced', 'married', 'frenzied', 'refined']	['much', 'uncomfortable', 'bitchy', 'feathery', 'about', 'childish', 'uneasy', 'afraid', 'fearful', 'dreamlike']	['unnerves', 'pore', 'millionth', 'yes', 'upwards', 'were', 'percentile', 'intersexes', 'planetesimals', 'eastwards']	['contemplative', 'tentative', 'amenable', 'companionable', 'constructive', 'decent', 'decadent', 'antisocial', 'prolific', 'demonic']	['cranky', 'brassy', 'shadowy', 'foreground', 'rotund', 'leathery', 'thick', 'sleeveless', 'colorful', 'pasty']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['grant', 'victor', 'presto', 'kingdom', 'agua', 'church', 'c', 'est', 'bureau']	['rent', '@', 'hangover', 'hmm', 'shhh', 'fanfare', 'matouchka', 'aqu', 'bueno']	['day', 'joe', 'pause', 'god', 'ave', 'lady', 'prince', 'uncle', 'auntie']	['bulbs', 'frisbee', 'hill', 'creek', 'pow', 'dragons', 'sun', 'tea', 'baba']	['melanoma', 'life', 'wart', 'diablo', 'qu', 'pero', 'act', 'out', 'sola']	['waters', 'es', 'workers', 'sabs', 'pues', 'yips', 'obits', 'dogs', 'dots']	['mission', 'orient', 'al', 'theory', 'stabilize', 'rules', 'quotations', 'i.', 'facts']	['barbed', 'ed', 'stalled', 'ted', 'met', 'crowed', 'called', 'drew', 'seed']
Adjectives	['decadent', 'ebullient', 'interested', 'pristine', 'ageless', 'trendy', 'unemployed', 'stately', 'rewarding', 'patronizing']	['ornery', 'foreground', 'rotund', 'lumpish', 'gaudy', 'housedress', 'feeble', 'upstage', 'tawdry', 'supine']	['rapt', 'oozy', 'pleased', 'bitchy', 'about', 'childish', 'wolfish', 'afraid', 'ticklish', 'arrogant']	['cranky', 'brassy', 'feathery', 'shadowy', 'ornate', 'leathery', 'gauzy', 'crunchy', 'thick', 'humid']	['much', 'contemplative', 'tentative', 'uncomfortable', 'obvious', 'endless', 'threatening', 'uneasy', 'incapable', 'companionable']	['ranging', 'bulk', 'thousandth', 'low-', 'smaller', 'upwards', 'were', 'intersexes', 'planetesimals', 'orthogonals']	['amenable', 'constructive', 'underlying', 'antisocial', 'prolific', 'destructive', 'subliminal', 'meaningful', 'incidental', 'discerning']	['unlocked', 'wrinkled', 'terrified', 'puzzled', 'dejected', 'unconvinced', 'frenzied', 'refined', 'amused', 'horrificed']

2000s

Word2vec (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['missy', 'jesus', 'mary', 'mo', 'sally', 'mark', 'simon', 'sam', 'sr']	['masta', 'preacha', 'ise', 'belmonti', 'yessah', 'fornicate', 'nossah', 'yessum', 'alsace']	['count', 'crawl', 'look', 'step', 'excuse', 'shoot', 'bet', 'switch', 'sleep']	['orleans', 'mississippi', 'tavern', 'den', 'sunday', 'germany', 'louisiana', 'river', 'towers']	['america', 'state', 'europe', 'court', 'courts', 'masters', 'mothers', 'neighbors', 'parts']	['bible', 'god', '@', 'details', 'control', 'i', 'authority', 'earth', 'act']	['watchman', 'salome', 'hellfire', 'satan', 'dem', 'mah', 'karla', 'muller', 'missus']	['liquor', 'a', 'wall', 'darkness', 'dance', 'bow', 'spider', 'hand', 'nose']

Adjectives	['pleased', 'impressed', 'upset', 'embarrassed', 'disappointed', 'predecessor', 'divorced', 'wow', 'ma'am']	['africanamerican', , 'ebey', 'ruint', 'pointin', 'miserable', 'marriageable', 'unable', 'distressing', 'utmost', 'womanly']	['pretend', 'refuse', 'swear', 'enjoy', 'toss', 'indicate', 'handed', 'reduce', 'endure']	['overgrown', 'crumbling', 'sprawling', 'nearest', 'deserted', 'wooded', 'isolated', 'parked', 'hurried', 'neighboring']	['expensive', 'involved', 'substantial', 'eligible', 'sufficient', 'numerous', 'extensive', 'administrative', , 'available', 'likely']	['afraid', 'overall', 'else', 'much', 'difficult', 'due', 'very', 'aware', 'interested', 'considerable', ,]	['menial', 'throaty', 'lusty', 'fervent', 'scarce', 'spiteful', 'excited', 'fearful', 'heartbroken', , 'kinda']	['comfortable', 'silky', 'dull', 'slight', 'strewn', 'identical', 'ruffled', 'motionless', 'colorful', 'below']
-------------------	---	--	--	---	---	--	--	--

Word2vec (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['masta', 'yessah', 'nossah', 'yessum', 'balize', 'lahble', 'balieve', 'bronck', 'broncks']	['watchman', 'den', 'hellfire', 'wall', 'flying', 'blackout', 'giggles', 'rushes', 'darkness']	['orleans', 'mississippi', 'tavern', 'sunday', 'germany', 'america', 'louisiana', 'river', 'watts']	['preacha', 'ise', 'belmonti', 'fornicate', 'dem', 'mah', 'alsace', 'henrik', 'ahse']	['jesus', 'bible', 'satan', 'god', 'teen', 'annerle', 'salzburg', 'empress', 'nannerle']	['liquor', 'pepsi', 'water', 'pizza', 'chile', 'hershey', 'flowers', 'oreos', 'deli']	['salome', 'missy', 'mary', 'mo', 'karla', 'sally', 'muller', 'missus', 'dey']	['state', 'a', '@', 'details', 'court', 'courts', 'control', 'space', 'authority']
Adjectives	['ebey', 'pug-nosed', 'forgedaboutit', , 'countriesand', 'subliminal', 'forgedaboutit', , 'onlyest', 'inner-peace', 'innerested', 'incested']	['throaty', 'geometric', 'intricate', 'teary', 'furry', 'outstretched', 'silky', 'frayed', 'drab', 'dull']	['umpteenth', 'quaint', 'luxurious', 'swanky', 'secluded', 'run-down', 'forested', 'opulent', 'populous', 'desolate']	['africanamerican', , 'ruint', 'pointin', 'hotheaded', 'outmoded', 'unmusical', 'cha-cha-cha', 'professional', 'ditsy', 'cookin']	['lusty', 'fervent', 'spiteful', 'excited', 'fearful', 'afraid', 'heartbroken', 'compulsive', 'sordid', 'affectionate']	['cheesy', 'plentiful', 'deviled', 'optional', 'stale', 'waxy', 'sugary', 'pungent', 'acrid', 'syrupy']	['kinda', 'incredulous', , 'tellin', 'sayin', 'doin', 'pissed', 'mousy', 'irate', 'hearin', 'goddamned', ,]	['menial', 'scarce', 'overall', 'redundant', 'acceptable', 'due', 'unnecessary', 'restrictive', 'expensive', 'considerable']

GloVe	c0	c1	c2	c3	c4	c5	c6	c7
Names	['missy', 'liquor', 'satan', 'sr', 'kings', 'rushes', 'vienna', 'inn', 'austria']	['watchman', 'masta', 'preacha', 'ise', 'belmonti', 'yessah', 'fornicate', 'nossah', 'mo']	['a', '@', 'i', 'day', 'man', 'look', 'time', 'us', 'may']	['mississippi', 'bible', 'germany', 'manhattan', 'tower', 'flying', 'amy', 'asia', 'nicky']	['salome', 'hellfire', 'dem', 'mah', 'karla', 'muller', 'missus', 'alsace', 'henrik']	['orleans', 'tavern', 'den', 'sally', 'louisiana', 'towers', 'empire', 'simon', 'cal']	['god', 'state', 'building', 'york', 'street', 'mom', 'father', 'house', 'control']	['jesus', 'mary', 'sunday', 'america', 'river', 'mark', 'park', 'sam', 'cont']
Adjectives	['fearful', 'kinda', 'intricate', 'outstretched', 'quaint', 'troubling', 'silky', 'arrogant']	['africanamerican', , 'throaty', 'scarce', 'ebey', 'ruint', 'pointin', 'amusing', 'furry', 'hesitant', 'phenomenal']	['much', 'very', 'on', 'through', 'around', 'before']	['excited', 'overall', 'comfortable', 'impressed', 'upset', 'unable', 'expensive', 'slight']	['menial', 'lusty', 'fervent', 'spiteful', 'heartbroken', 'incredulous', 'compulsive', 'geometric']	['acceptable', 'frightened', 'miserable', 'unhappy', 'annoyed', 'respected', 'fortunate', 'suitable', 'alike', 'dull']	['else', 'likely', 'done', 'behind', 'ago', 'suddenly', 'along']	['afraid', 'worried', 'pleased', 'difficult', 'due', 'surprised', 'aware', 'interested', 'earlier', 'familiar']

	'kindly', 'unnecessary']			'uncomfortable', 'embarrassed']	'redundant', 'umpteenth']			
--	-----------------------------	--	--	------------------------------------	------------------------------	--	--	--

fastText (CBOW)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['building', 'flying', 'king', 'ring', 'riding', 'dang', 'bombing', 'pal', 'bunting']	['mary', 'dem', 'tem', 'state', 'cal', 'cont', 'count', 'court', 'control']	['river', 'park', 'a', 'street', 'tower', 'wall', 'beach', 'narrows', 'blackout']	['tavern', 'missus', 'dey', 'mark', 'empire', 'magnesia', 'biographer', 'seventeen', 'empress']	['missy', 'bible', 'mah', 'mo', 'god', 'sally', 'mom', '@', 'mama']	['den', 'teen', 'ned', 'en', 'grant', 'ed', 'ted', 'gran', 'reed']	['watchman', 'liquor', 'jesus', 'sunday', 'hellfire', 'pueblo', 'archbishop', 'sixteen', 'von']	['watts', 'towers', 'kings', 'giggles', 'details', 'courts', 'masters', 'saints', 'neighbors']
Adjectives	['amusing', 'enchanted', 'troubling', 'upsetting', 'distressing', 'exhausting', 'upcoming', 'fascinating', 'disgusting', 'exciting']	['menial', 'intricate', 'affectionate', 'difficult', 'phenomenal', 'presentable', 'personable', 'conspiratorial', 'restrictive', 'expensive']	['throaty', 'silky', 'lowly', 'drab', 'dull', 'slight', 'windowless', 'strewn', 'cheesy', 'rid']	['fervent', 'scarce', 'incredulous', 'compulsive', 'geometric', 'overall', 'redundant', 'acceptable', 'teary', 'quaint']	['spiteful', 'fearful', 'afraid', 'heartbroken', 'kinda', 'sordid', 'else', 'much', 'hesitant', 'comfortable']	['excited', 'worried', 'outstretched', 'hotheaded', 'pleased', 'frightened', 'fatigued', 'astounded', 'annoyed', 'impressed']	['lusty', 'umpteenth', 'furry', 'arrogant', 'fussy', 'cha- cha-cha', 'clunky', 'swanky', 'outspoken', 'carnivorous']	['costly', 'thousandth', 'millionth', 'upwards', 'indoors', 'denotes', 'somewheres', 'earthmovers', 'onwards']

fastText (Skip-gram)	c0	c1	c2	c3	c4	c5	c6	c7
Names	['liquor', 'den', 'mah', 'mo', 'sally', 'river', 'tem', 'tower', 'wall']	['empire', 'details', 'control', 'university', 'authority', 'temptations', 'judgment', 'dimension', 'persuaders']	['flying', 'king', 'ned', 'serpent', 'ring', 'ed', 'riding', 'dang', 'drew']	['state', 'biographer', 'seventeen', 'archbishop', 'von', 'tour', 'court', 'courts', 'councillor']	['watts', 'towers', 'kings', 'masters', 'saints', 'mothers', 'neighbors', 'eighties', 'parts']	['watchman', 'missy', 'tavern', 'hellfire', 'god', 'dey', 'a', 'street', 'cont']	['jesus', 'bible', 'mary', 'dem', 'missus', 'magnesia', 'cal', 'grace', 'empress']	['sunday', 'mark', 'building', 'park', 'pueblo', 'blackout', '@', 'june', 'winter']
Adjectives	['throaty', 'lusty', 'furry', 'silky', 'stuffy', 'lowly', 'drab', 'dull', 'clunky', 'windowless']	['overall', 'acceptable', 'phenomenal', 'due', 'restrictive', 'educated', 'hypocritical', 'understandable', 'abusive', 'psychopathic']	['outstretched', 'hotheaded', 'enchanted', 'frightened', 'astounded', 'impressed', 'infatuated', 'besotted', 'wretched', 'frayed']	['umpteenth', 'respected', 'secretive', 'conspiratorial', 'earlier', 'outspoken', 'upcoming', 'professorial', 'plannin', 'delinquent']	['thousandth', 'auto-parts', 'migr', 'upwards', 'indoors', 'scattershot', 'somewheres', 'thousandfold', 'dot-coms', 'everywhere']	['excited', 'afraid', 'heartbroken', 'kinda', 'teary', 'worried', 'else', 'pleased', 'hesitant', 'annoyed']	['menial', 'fervent', 'scarce', 'spiteful', 'fearful', 'incredulous', 'compulsive', 'geometric', 'redundant', 'intricate']	['quaint', 'expensive', 'luxurious', 'cha-cha-cha', 'carnivorous', 'priceless', 'offhand', 'genteel', 'straightforward', 'big-']

APPENDIX B. JACCARD SIMILARITY BETWEEN WORD EMBEDDING CLUSTERS

1810	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.39	0.48	0.37	0.36
Word2vec (Skip-gram)	0.39	1.00	0.32	0.47	0.47
GloVe	0.48	0.32	1.00	0.33	0.29
fastText (CBOW)	0.37	0.47	0.33	1.00	0.57
fastText (Skip-gram)	0.36	0.47	0.29	0.57	1.00

1820	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.36	0.44	0.34	0.33
Word2vec (Skip-gram)	0.36	1.00	0.33	0.45	0.46
GloVe	0.44	0.33	1.00	0.33	0.32
fastText (CBOW)	0.34	0.45	0.33	1.00	0.63
fastText (Skip-gram)	0.33	0.46	0.32	0.63	1.00

1830	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.41	0.44	0.35	0.39
Word2vec (Skip-gram)	0.41	1.00	0.34	0.46	0.50
GloVe	0.44	0.34	1.00	0.31	0.35
fastText (CBOW)	0.35	0.46	0.31	1.00	0.58
fastText (Skip-gram)	0.39	0.50	0.35	0.58	1.00

1840	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.40	0.40	0.35	0.38
Word2vec (Skip-gram)	0.40	1.00	0.34	0.47	0.52
GloVe	0.40	0.34	1.00	0.32	0.36
fastText (CBOW)	0.35	0.47	0.32	1.00	0.60
fastText (Skip-gram)	0.38	0.52	0.36	0.60	1.00

1850	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.32	0.44	0.31	0.35
Word2vec (Skip-gram)	0.32	1.00	0.38	0.48	0.53
GloVe	0.44	0.38	1.00	0.37	0.43
fastText (CBOW)	0.31	0.48	0.37	1.00	0.64
fastText (Skip-gram)	0.35	0.53	0.43	0.64	1.00

1860	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.38	0.47	0.34	0.40
Word2vec (Skip-gram)	0.38	1.00	0.34	0.45	0.52
GloVe	0.47	0.34	1.00	0.33	0.39
fastText (CBOW)	0.34	0.45	0.33	0.91	0.65
fastText (Skip-gram)	0.40	0.52	0.39	0.65	1.09

1870	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.39	0.46	0.33	0.39
Word2vec (Skip-gram)	0.39	1.00	0.35	0.41	0.48
GloVe	0.46	0.35	1.00	0.32	0.38
fastText (CBOW)	0.33	0.41	0.32	1.00	0.55
fastText (Skip-gram)	0.39	0.48	0.38	0.55	1.00

1880	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.35	0.43	0.30	0.37
Word2vec (Skip-gram)	0.35	1.00	0.32	0.44	0.51
GloVe	0.43	0.32	1.00	0.30	0.36
fastText (CBOW)	0.30	0.44	0.30	1.00	0.61
fastText (Skip-gram)	0.37	0.51	0.36	0.61	1.00

1890	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.37	0.42	0.33	0.40
Word2vec (Skip-gram)	0.37	1.00	0.36	0.46	0.52
GloVe	0.42	0.36	1.00	0.35	0.43
fastText (CBOW)	0.33	0.46	0.35	0.91	0.63
fastText (Skip-gram)	0.40	0.52	0.43	0.63	1.09

1900	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.40	0.45	0.34	0.41
Word2vec (Skip-gram)	0.40	1.00	0.37	0.46	0.51
GloVe	0.45	0.37	1.00	0.35	0.42
fastText (CBOW)	0.34	0.46	0.35	1.00	0.65
fastText (Skip-gram)	0.41	0.51	0.42	0.65	1.00

1910	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.40	0.43	0.34	0.42
Word2vec (Skip-gram)	0.40	1.00	0.37	0.45	0.51
GloVe	0.43	0.37	1.00	0.37	0.44
fastText (CBOW)	0.34	0.45	0.37	1.00	0.63
fastText (Skip-gram)	0.42	0.51	0.44	0.63	1.00

1920	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.37	0.42	0.33	0.39
Word2vec (Skip-gram)	0.37	1.00	0.38	0.47	0.52
GloVe	0.42	0.38	1.00	0.39	0.44
fastText (CBOW)	0.33	0.47	0.39	1.00	0.70
fastText (Skip-gram)	0.39	0.52	0.44	0.70	1.00

1930	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.42	0.44	0.36	0.41
Word2vec (Skip-gram)	0.42	1.00	0.39	0.47	0.49
GloVe	0.44	0.39	1.00	0.39	0.44
fastText (CBOW)	0.36	0.47	0.39	1.00	0.64
fastText (Skip-gram)	0.41	0.49	0.44	0.64	1.00

1940	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.37	0.41	0.35	0.39
Word2vec (Skip-gram)	0.37	1.00	0.35	0.49	0.52
GloVe	0.41	0.35	1.00	0.37	0.42
fastText (CBOW)	0.35	0.49	0.37	1.00	0.66
fastText (Skip-gram)	0.39	0.52	0.42	0.66	1.00

1950	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.38	0.43	0.32	0.38
Word2vec (Skip-gram)	0.38	1.00	0.41	0.47	0.52
GloVe	0.43	0.41	1.00	0.37	0.44
fastText (CBOW)	0.32	0.47	0.37	1.00	0.58
fastText (Skip-gram)	0.38	0.52	0.44	0.58	1.00

1960	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.34	0.44	0.31	0.40
Word2vec (Skip-gram)	0.34	1.00	0.38	0.46	0.52
GloVe	0.44	0.38	1.00	0.36	0.46
fastText (CBOW)	0.31	0.46	0.36	1.00	0.57
fastText (Skip-gram)	0.40	0.52	0.46	0.57	1.00

1970	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.34	0.45	0.29	0.34
Word2vec (Skip-gram)	0.34	1.00	0.31	0.44	0.51
GloVe	0.45	0.31	1.00	0.30	0.35
fastText (CBOW)	0.29	0.44	0.30	1.00	0.58
fastText (Skip-gram)	0.34	0.51	0.35	0.58	1.00

1980	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.41	0.45	0.35	0.42
Word2vec (Skip-gram)	0.41	1.00	0.37	0.42	0.47
GloVe	0.45	0.37	1.00	0.37	0.46
fastText (CBOW)	0.35	0.42	0.37	1.00	0.57
fastText (Skip-gram)	0.42	0.47	0.46	0.57	1.00

1990	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.34	0.43	0.31	0.41
Word2vec (Skip-gram)	0.34	1.00	0.37	0.46	0.51
GloVe	0.43	0.37	1.00	0.37	0.47
fastText (CBOW)	0.31	0.46	0.37	1.00	0.60
fastText (Skip-gram)	0.41	0.51	0.47	0.60	1.00

2000	Word2vec (CBOW)	Word2vec (Skip-gram)	GloVe	fastText (CBOW)	fastText (Skip-gram)
Word2vec (CBOW)	1.00	0.39	0.45	0.39	0.43
Word2vec (Skip-gram)	0.39	1.00	0.38	0.47	0.49
GloVe	0.45	0.38	1.00	0.44	0.48
fastText (CBOW)	0.39	0.47	0.44	1.00	0.74
fastText (Skip-gram)	0.43	0.49	0.48	0.74	1.00