# Exploring the Role of Gender in 19th Century Fiction Through the Lens of Word Embeddings

Siobhán Grayson[1], Maria Mulvany[2], Karen Wade[2], Gerardine Meaney[2], and Derek Greene[1]

[1] School of Computer Science, University College Dublin, Ireland
{siobhan.grayson, derek.greene}@insight-centre.org
[2] Humanities Institute, University College Dublin, Ireland
{maria.mulvany, karen.wade, gerardine.meaney}@ucd.ie

**Abstract.** Within the last decade, substantial advances have been made in the field of computational linguistics, due in part to the evolution of word embedding algorithms inspired by neural network models. These algorithms attempt to derive a set of vectors which represent the vocabulary of a textual corpus in a new embedded space. This new representation can then be used to measure the underlying similarity between words. In this paper, we explore the role an author's gender may play in the selection of words that they choose to construct their narratives. Using a curated corpus of forty-eight 19th century novels, we generate, visualise, and investigate word embedding representations using a list of gender-encoded words. This allows us to explore the different ways in which male and female authors of this corpus use terms relating to contemporary understandings of gender and gender roles.

## 1 Introduction

In the fields of natural language processing and text mining, the study of word co-occurrences has often been used to identify the linkages between words in unstructured texts. The motivation for this type of analysis comes from the distributional hypothesis in linguistics, which states that "a word is characterised by the company it keeps" [4]. One of the most popular approaches in the literature has been *word2vec* [10], which uses a two-layer neural network model to capture word contexts in a corpus, translating words into $d$-dimensional *word vectors*. This allows for the detection of contextually similar words without human intervention, as vectors for words with similar semantic meanings tend to be located close to one another. One interesting corollary to this is that biases such as gender stereotypes that may be implicitly present within a corpora, can be identified and studied from a quantitative perspective [3]. Such insights are beneficial to wide range of fields, including humanities, where an increasing number of scholars are seeking to complement their literary research by incorporating computational techniques to provide alternative perspectives [7].

This particularly benefits scholars who are interested in 'distant reading' [11], the practice of understanding literature from a macro-level viewpoint, as opposed to exclusively from a traditional micro-level 'close reading' standpoint. So far, a number of different computational methods have been applied to quantitatively study literature from a macro perspective. Jockers and Mimno [8] apply topic modelling to a large corpus of 19th-century fiction to identify broad themes. Whilst Reagan *et al.* [12] use sentiment analysis to understand the emotional arcs of 1,700 works of fiction from Project Gutenberg. More recently, Grayson *et al.* [5] applied word embeddings to explore 19th century fiction and investigate differences in characterisation between novels. While Heuser[2] analyses word associations produced by a *word2vec* model built on 18th-century texts, and Cherny[3] creates a visualisation of the nouns appearing in Jane Austen's *Pride and Prejudice*, generated using *word2vec* and the t-SNE visualisation method.

The most similar work to this paper is perhaps that of Schmidt [13], who uses embeddings to identify gender bias present within 'Rate My Professors' reviews and then proposes a *vector rejection* method for de-biasing embeddings by eliminating gender effects. However, unlike Schmidt, we do not seek to eliminate gender bias. Here we analyse word embeddings generated using a curated corpus of forty-eight British and Irish 19th century novels that have been manually annotated to include definitive character names[1]. We focus on uncovering the different contexts in which female and male authors of the 19th century engage with gender specific words, by compiling a list of gender-encoded unigrams, such as 'she' and 'he', and then annotating each of their occurrences within our corpus to reflect the author's gender of the text they appear in ('she_female', 'he_female'). We subsequently find differences which tally with those identified previously [1], where pronouns and nouns appear in different semantic spaces, depending on the gender of the author.

## 2 Methods

In this paper we consider a collection of forty-eight novels from twenty-nine 19th century novelists sourced from Project Gutenberg, summarised by author gender in Table 1. Initial data preparation involves the manual annotation of the novels, where literary scholars identify all character references in the text of each novel as described in [5]. The corpus was then further annotated using a list of gender encoded unigrams, see Fig. 1(a), where each of their occurrences within our corpus was labelled to reflect the author's gender of the text they appear in. Afterwards, part-of-speech tagging (POS tagging) was applied using the Natural Language Toolkit (NLTK) [2] PerceptronTagger implementation. For the purposes of converting our textual datasets into vector word embeddings, we employ a skipgram word2vec model [10].

Based on [5], word embeddings were generated using a skipgram model with 300 dimensions, a context window size of 5 words, and a minimum word fre-

[2] `http://ryanheuser.org/word-vectors-1`

[3] `http://www.ghostweather.com/files/word2vecpride`

[1] The annotated texts were created as part of the "Nation, Gender, Genre" project. See `http://www.nggprojectucd.ie`

Table 1: Summary of the corpus used in this work, by author gender.

| Gender | *#Authors* | *#Novels* | *#Characters* | *#Chapters* | *#Sentences* | *#Words* | *%Words* |
|---|---|---|---|---|---|---|---|
| Female | 11 | 22 | 4005 | 816 | 111,102 | 2,707,884 | 46% |
| Male | 18 | 26 | 6436 | 983 | 136,023 | 3,130,090 | 54% |
| Total | 29 | 48 | 10,441 | 1,799 | 247,125 | 5,837,974 | |

quency of 50. All other parameters were left at their default settings. We then visualised the resulting embeddings by reducing the dimension of each vector into a 2D space using t-Distributed Stochastic Neighbour Embedding (t-SNE) [9]. Finally, to analyse the semantic differences in how female and male authors incorporated our list of gender encoded words, we computed the cosine similarity between each of the resulting *female* and *male* labelled word embeddings to measure how similarly these words are used by authors of different genders.

## 3 Results

The word frequency of the initial list of gender-encoded words is displayed in Fig. 1 (a), where bar lengths correspond to log frequency values, while the actual word frequency is displayed within the bars. The top four words are pronouns {he, her, she, him} where 'he' is the most frequently used word by both female and male authors, with male authors using 'he' almost double the number of times they use the second most frequent word 'she'. As described in Section 3, a minimum word frequency 50 was applied when training the word2vec model. Therefore, words highlighted in yellow do not appear in our final embeddings, as either one or both genders did not use these words more than 50 times within our corpus. In Fig.1 (b), the resulting cosine similarity of the remaining female and male annotated embeddings are displayed: higher scores equate to greater semantic similarity whilst lower scores indicate lower semantic similarity. In this case, 'fellow' is the word that appears to be used in the most semantically similar contexts for both female and male authors, while 'husband' appears to be used in the most semantically dissimilar contexts by both genders.

As well as calculating the cosine similarity between gender annotated embeddings, we have visualised all embeddings in Fig. 2. Gender-encoded unigrams by female authors are depicted as large, pink circles while the corresponding male authored unigrams are depicted as large, grey circles. In particular, we found gender-encoded embeddings to occupy four different spaces within our embeddings projection. These spaces have been annotated A-D in Fig. 2. Group A consists of both female- and male-authored plural nouns {*fellows, women, men,..*} from our gender-encoded list, see Fig. 1.(b), nested within a pocket of past-participles verbs. However, no family related nouns such as {*daughters, sisters, brothers*} by female authors are contained despite the presence of their male-authored counterparts. Group B is the largest of our clusters and consists of singular gender-encoded nouns by both genders surrounded by nouns refer-

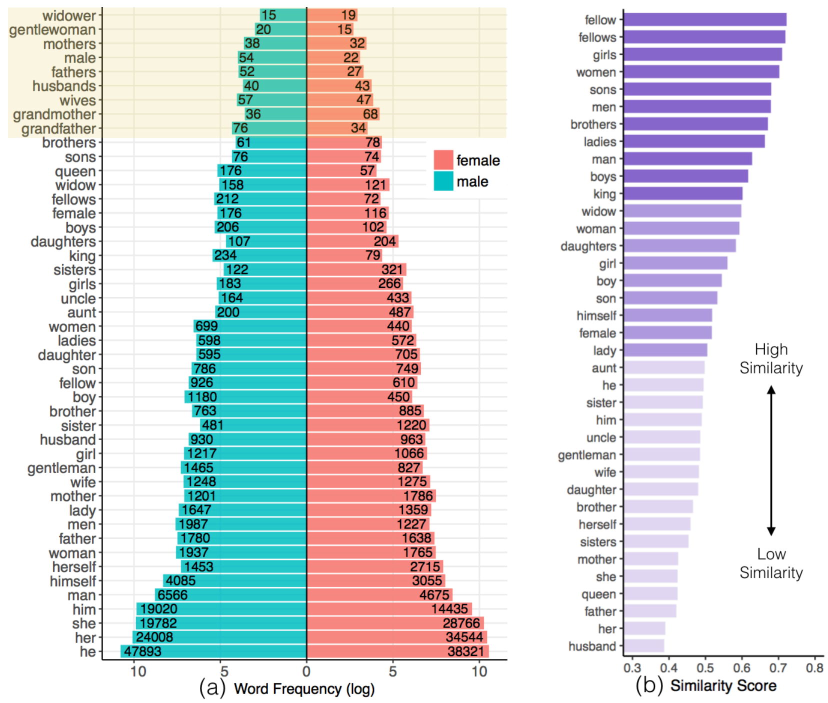


Fig. 1: (a) Word frequencies for our initial list of gender-encoded words. (b) The cosine similarity scores between female and male authored words in our gender-encoded list.

ring primarily to (typically male) occupations, such as "priest", "clerk", "magistrate", and "farmer". However, it also contains all the male-authored pronouns within our list, again see Fig. 1(b), but only one female authored pronoun, "himself". The rest of our female authored pronouns are found within Group D, next to a mixture of past-participles (blue) and past verbs (purple). This provides and interesting counterpoint to Argamon *et al.* [1] who found differences in how women and men used words, particularly personal pronouns. Meanwhile, Group C consists of family related nouns (singular and plural) by only female authors, nested within a cluster of characters predominately from Jane Austen's novels.

Finally, we have analysed the nearest neighbours for each gender-encoded unigram. The differences between male and female authors' use of the word "her" are particularly striking. In works by female authors, "her" is frequently found alongside terms pertaining to emotional experiences, including "shrinking", "sobs", "trembling", and "flutter". By contrast, the pronoun's nearest neighbours in male-authored texts include "she", "him", and "his". Further details of the nearest neighbouring words after filtering out character names for a subset of our gender-encoded unigrams are included in Table 2. Where we observe similar behaviour in how the pronoun "he" is used differently depending on the gender of the author. Again this tallies with what has previously been

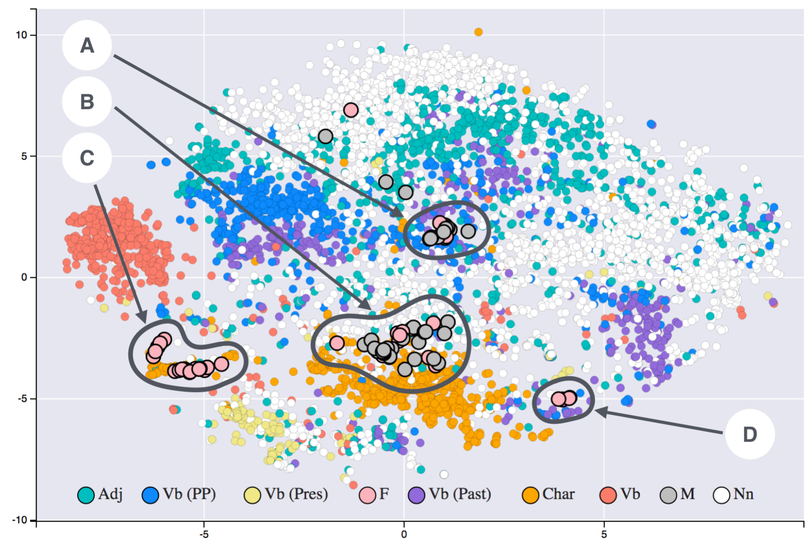


Fig. 2: Embeddings generated from our entire corpus visualised using t-SNE and coloured according to their grammatical class. Adjective: Green, Verb (Past Participle): Blue, Verb (Present): Yellow, Female: Large pink, Verb (Past): Purple, Character: Orange nodes, Verb: Red, Male word: Large grey, Noun: White.

Table 2: Selected gendered words and their nearest neighbours where superscripts denote that a word is apart of our gender-encoded unigrams list and indicates whether it is a female ($f$) or male ($m$) authored embedding.

| Word | Gender | 8 Nearest Neighbours |
|---|---|---|
| He | F | she$^f$, him$^f$, her$^f$, he$^m$, himself$^f$, vaguely, nervously, trembling |
| | M | she$^m$, him$^m$, himself$^m$, his, he$^f$, her$^m$, it, that |
| Lady | F | gentleman$^f$, woman$^f$, girl$^f$, ladies$^f$, heiress, lady$^m$, widow$^f$, maid |
| | M | woman$^m$, gentleman$^m$, girl$^m$, aunt$^m$, widow$^m$, major, maid, friend |
| Gentleman | F | lady$^f$, man$^f$, farmer, clergyman, bachelor, barrister, nobleman, lawyer |
| | M | soldier, man$^m$, lady$^m$, officer, magistrate, farmer, nobleman, colonel |

found by Argamon *et al.* [1] with respect to pronouns. The second observation is that both female and male authors tend to use the word “gentleman” in similar spaces as occupation, whilst we see the word “girl” make an appearance in both gendered neighbour lists for “lady”, although the converse is not true for “boy” which is absent from both lists for “gentleman”.

## 4 Conclusion

In this paper, we explored the differences between word use by male and female authors in a corpus of 19th century novels. Having generated, visualised and analysed word embedding representations using a list of gender-encoded word pairs, we found that there are differences in the ways in which the male and female authors of this corpus use terms relating to contemporary understandings of gender and gender roles (such as "she", "lady", "gentleman" and occupations/professions). Our results correspond with those of Argamon *et al.* [1], who identified significant differences in the use of personal pronouns in the writing of men and women. Although identifying the meaning of these gendered differences is beyond the scope of this preliminary survey, our analysis of word embeddings (as shown in Fig. 2) shows marked differences in the use of gendered pronouns by male and female authors. In future work, we hope to extend the size of our corpus to allow for diachronic word embedding analysis [6], in order to explore potential differences arising as a result of the era in which a novel was written and to clarify how this interacts with the gender differences we have identified.

**Acknowledgments.** This research was partly supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289, in collaboration with the Nation, Genre and Gender project funded by the Irish Research Council.

## References

1. Argamon, S., Koppel, M., Fine, J., Shimoni, A.R.: Gender, genre, and writing style in formal written texts. TEXT 23, 321–346 (2003)
2. Bird, S., Klein, E., Loper, E.: Natural language processing with Python. " O'Reilly Media, Inc." (2009)
3. Bolukbasi, T., Chang, K.W., Zou, J.Y., Saligrama, V., Kalai, A.T.: Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In: Advances in Neural Information Processing Systems. pp. 4349–4357 (2016)
4. Firth, J.R.: A synopsis of linguistic theory 1930-55. 1952-59, 1–32 (1957)
5. Grayson, S., Mulvany, M., Wade, K., Meaney, G., Greene, D.: Novel2vec: Characterising 19th century fiction via word embeddings. In: Proc. 24 Irish AICS (2016)
6. Hamilton, W.L., Leskovec, J., Jurafsky, D.: Diachronic word embeddings reveal statistical laws of semantic change. In: Proc. of 54th ACL (2016)
7. Jockers, M.L.: Macroanalysis: Digital methods and literary history. University of Illinois Press (2013)
8. Jockers, M.L., Mimno, D.: Significant themes in 19th-century literature. Poetics 41(6), 750–769 (2013)
9. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research 9(Nov), 2579–2605 (2008)
10. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: Workshop Proc. ICLR (2013)
11. Moretti, F.: Network Theory, Plot Analysis. New Left Review 68, 80–102 (2011)
12. Reagan, A.J., Mitchell, L., Kiley, D., Danforth, C.M., Sheridan Dodds, P.: The emotional arcs of stories are dominated by six basic shapes. ArXiv e-prints (2016)
13. Schmidt, B.: Rejecting the gender binary: a vector-space operation. (2015), `http://bookworm.benschmidt.org/posts/2015-10-30-rejecting-the-gender-binary.html`