Social Bias Detection in Literary Texts, Leveraging Graph Representations

Master's Thesis Proposal

Furkan Akkurt
Advised by Susan M Üsküdarlı

Boğaziçi University
Department of Computer Engineering
Istanbul, Turkey

January 23, 2024

Contents

Abstract

Linguistic expressions betray biases even if their producers are not cognizant of them. Let alone malicious actors, even benign or ordinary people may be perpetuating biased language. These expressions are also increasingly used to train machine learning models for real-life applications. Many studies tackled the problem of detecting social biases (e.g. racial, gender). These studies mostly focus on language in use (e.g. social media, news), and do not consider pinpointing sections of language outputs with bias. In this study, we aim to work on literary texts, which abound with social biases, and are not studied as much as texts deemed as language in use. Since literary works tend to be longer, they provide an opportunity to study social biases in more complex linguistic environments. Our task is to pinpoint various types of social bias, such as racial, gender and class-based ones.

Graphs are applicable for representing linguistic expressions as they enable relating various linguistic structures, they are used extensively in linguistics and many relevant machine learning studies make use of them. Graph neural networks are able to learn rich latent representations of graph-structured data for various tasks. As such, we plan to use graphs to learn representations of expressions that are significant for bias detection. We plan, also, to utilize existent knowledge graphs (i.e. databases storing graph-structured knowledge) as they hold significant information on semantics of expressions. We plan to construct a general method for bias detection, which may be used in other domains than the literary world. Using this method, social applications that are able to detect biases before they are published or rate documents based on their biases may be developed.

Definitions

For the purposes of this proposal, let's define the following terms so as to reduce ambiguity:

- *Word*: a sequence of characters (i.e. letters), in many orthographies separated by spaces. Example: "beautiful". Let c be a character and w a word. $w = c_1, c_2, \ldots, c_n$ where c_i is the i'th character of the word. $|w| \geq 1$. A one-character word example is "a" in English.
- *Sentence*: a sequence of words, usually ending with a punctuation mark. Example [1]: "No such thing as a life that's better than yours." Let m be a punctuation mark. Let s be a sentence, an ordered list of n words with an ending punctuation mark m: $s = w_1, w_2, \ldots, w_n, m$. $|s| \geq 1$.
- *Phrase*: a sequence of words (e.g. 'telltale sign'), usually separated by spaces. A phrase may be made up of a single word. Let p be a phrase. $p = w_1, w_2, \ldots, w_n$ where w_i is the i'th word of the phrase. $|p| \geq 1$.
- *Document*: a sequence of sentences (e.g. a book, a report). Let d be a document. $d = s_1, s_2, \ldots, s_n$ where s_i is the i'th sentence of the document. $|d| \geq 1$.
- *Literary text*: a document of literature, such as a novel or a poem. They are characterized by their high quality of form. Example: "The Great Gatsby" [2] by F. Scott Fitzgerald.
- *Agent*: a person or a group of people. Let a be an agent. $a \in A$ where A is the set of all agents in a document. For example, in Aesop's fable "The Fox & the Grapes" [3], A includes only the fox.
- *Graph*: a collection of nodes and edges. Let G be a graph, $G = (V, E)$ where V is the set of nodes and E is the set of edges.
- *Node*: an element of a graph, alternatively called a vertex. Let v be a node, $v \in V$.

- *Node features or embeddings*: a set of numerical features associated with a node. They may be pre-computed or learned. For example, pre-computed node features may include the part-of-speech tag of a word.
- *Edge*: a relationship between two nodes of a graph. Let e be an edge, $e \in E$ where $e = (v_1, v_2)$.
- *Edge features or embeddings*: a set of numerical features associated with an edge. They may be pre-computed or learned. For example, pre-computed edge features may include the dependency relation between two words.

1 Introduction

Speakers reflect their social biases that they hold, whether conscious or otherwise, through their speech or writings. Some types of these social biases can be said to be based on such social constructs as "race" or "gender". For an example of racial bias, refer to the excerpt from the classic novel by Harper Lee, "To Kill a Mockingbird" [4]: "A small boy clutching a *Negro* woman's hand walked toward us. He looked *all Negro* to me; he was rich *chocolate* with *flaring nostrils* and beautiful teeth." [1]. In this example, the character is shown to hold physical stereotypes regarding people of dark complexion.

Literary texts provide an ample source for analyses of these biases. They include narratives with plot lines, characters and actions they perform. Bias not only exists within actions, whether behavioral or by speech, of characters, but also in the language of the author. Considering this, some linguistic units, unbeknownst to the characters, can also be said to be biased. For example, adjectives an author uses to describe only one gender may surface as the source of the bias in a text. In fact, bias in a text is solely created and owned by its author, not a specific group of people. Analyzing lots of texts from a specific group of people may light up some areas of bias where the group is consistent on, reflecting the culture.

In this study, we hope to develop a method that we would be able to utilize to detect various types of social bias in literary texts. Graphs are used extensively in linguistics to represent language syntactically (e.g. syntax trees [5]) or semantically (e.g. semantic networks [6]), since they allow to capture various relationships between linguistic elements, which is why we also would like to utilize them in this study. Representing literary writings as graphs, we will try to learn representations from linguistic signals that are significant for bias detection. Even though we would like to focus our experiments on literary texts and analyze the literature domain specifically, we plan to construct the eventual architecture to be highly general. It is hoped that it could theoretically be applied to any type of text, not just

[1] Some parts are in italics for emphasis.

literary; to any type of bias: gender, racial, or even ornithocentric (related to birds); and also to any language, not just English. We acknowledge that it is possible that the proposed method may not be able to be used for low-resource languages (constituting most of the world's languages) as some parts of the method may not be readily available for them.

The rest of the proposal is organized as follows: Section 2 articulates the research question that we would like to answer during the present study. Section 3 provides a background on social bias: how it's defined and some examples from fictional works. Section 4 reviews the literature on bias detection. Section 5 describes the data that we plan to use. Section 6 presents the research methodology that we plan to follow; Section 7 reports the expected outcomes; and Section 8 tabulates a tentative timeline for the thesis. Finally, Section 9 concludes the proposal.

2 Research Question

Biased language perpetuates biased behavior in people. Additionally, this type of language is being fed into more and more machine learning models to be used in various applications [7], such as the recently unveiled Gemini, a family of large language models developed by Google [8]. In order to reduce biased treatment, first, one needs to detect biased usage of language and inform speakers accordingly. The main problem this thesis will try to tackle is the local detection of various types of social bias, whether implicit or explicit, existent specifically in literary texts. Given a document, the eventual proposed method should be able to show *what* kinds of social bias exist and *where* in the document. The location should be as specific as possible, e.g. a sentence or a phrase, labeled with the kind of bias (e.g. class-based).

3 Background on Social Bias

Social bias can be defined as the **effect of social category cues** (e.g. ethnic cues) on **behavioral** responses [9]. In other words, some signals based on different social categories (e.g. gender) can make people act differently, producing social bias. These actions can be benign or harmful [10], in a continuum. They can be explicit, easier to detect, or implicit, not expressed directly. In this study, we will concern ourselves with objective biases. By the word "objective", the quantitative distributions of linguistic signals, complex or otherwise, are meant. The following subsections define and exemplify types of social biases that we will consider in this study.

3.1 Gender

Gender bias is behavior that shows *favoritism* toward one gender over another. Gender is defined as socially constructed expectations and roles for women and men, girls and boys. An example of gender bias from the article 'Bias, Fiction, and the Negro' [11], *italics* for emphasis: "When a "white" novelist handles the problem, *he* may, in the beginning, present the Negro characters in a favorable light: ..." Using "he" assumes a general novelist to be male, so the example is biased regarding males.

3.2 Racial & Ethnic

Racial bias refers to biased treatment based on race. It often shows up linguistically through stereotypes or slurs [12]. In literature, it can either be outright (explicit) or subtle (implicit), influencing character attributes, dialogue, and narrative elements. An example [13] from "Of Mice and Men" (1937), where Crooks, the only black man in the context, says (*italics* for emphasis): "They play cards in there, but I can't play because I'm *black*." In this example, the character reports they can't do a specific act because they're of one race; this utterance wouldn't come from the other races (e.g. white) in the book. This is an example of racial power structure showing up as a bias. Another example from the fictional work 'Gone With The Wind' [14], *italics* for emphasis: "Mammy *was black, but* her code of conduct and her sense of pride were as high as or higher than those of her owners." In this example, the implication is that being black is in the way of someone's code of conduct being high, a *racial* bias.

While race is generally determined by physical characteristics, ethnicity refers to the cultural elements that make a group of people distinct. Ethnic bias is one where these cultural elements are the source of bias rather than physical features, as happens with racial bias.

3.3 Social class

Social class bias can be defined as biased behavior towards an individual or a group of people within a certain social class, generally correlated with their socioeconomic status. An example from the book 'Untouchable' [15], *italics* for emphasis: "Now Sohini being of the lowest caste among the out-castes would *naturally be looked down upon* by Gulabo." Looking down upon somebody *only* based on caste (a system of social stratification) is a class bias. The caste system of the book's context highly correlates with the socioeconomic categories the people find themselves in.

3.4 Intersectional

Some biases occur together, amplifying the harm caused. Intersectionality refers to how various types of biases intersect a person's multiple, sometimes overlapping, identities [16]. Intersecting biases can be said to be more harmful than the sum of the individual biases they're made up of. In 'Demarginalizing the Intersection of Race and Sex', Crenshaw exemplifies black women's experience where they're excluded from feminist and antiracist theories due to the theories being grounded in single axes. In the gender axis, feminist theories focus on 'white' women; while in the race axis, antiracist theories focus on black 'males'; essentially excluding 'black women' in both theories. This is an example of intersectional bias within the general architecture of the mentioned theories. Since theories are constructed, explained, and discussed using language, the language used in the theories will have to be biased, intersectionally.

4 Literature Review

Many natural language processing applications revolve around detecting biased language in texts, be it hate speech [17] or toxic language biased towards a specific group. The task is usually text classification [18], whether binary or multi-class, where the models predict whether a given text is biased or not, if the former; or biased to what degree (discrete) or not at all, if the latter. Traditional approaches used Bayesian models, lexicons [19] or manual features [20]. Deep learning methods improved on the previous results by using word embeddings [21] and neural networks [22]. These methods usually require supervision of annotated data, and human annotation is prone to bias itself. Also, if annotations are needed to develop a model, for each domain and language, a new set of annotations is needed, which is costly and time-consuming to acquire.

Most works focus on detecting explicit biases in texts of social media or news pieces, language in use in an immediate context. Implicit biases also exist in these texts but are harder to find. Fictional works present themselves as a good resource for studying implicit biases as they have more complex structures and just as much of representational power, if not higher. Even though there are lots of works on bias in the literature from a social science standpoint [23], there are not many computational works that focus on literary texts.

One work that does focus on literary texts analyzes gender bias in children's fairy tales [23]. Their proposed pipeline constructs a temporal event chain of a given story with character attributes. Here, temporal event chains are sequences of acts ordered by time. They use odds ratios for their analysis, reporting, for example, predominantly which types of actions each gender takes (e.g. cry, conquer).

5 Data

Texts of focus in this thesis will be literary ones. Literary texts differ from texts in various social media in that they're more formal and structured; and from news articles in that they're longer and have linguistically more complex sentences. Since the thesis will focus on literary texts, the data will be a collection of books. The books will be mostly in English, although not exclusively, from the public domain, meaning only books not under copyright currently. For example, in the US, as of January 1, 2024, all works published in 1929 or earlier are in the public domain [24]. These books are easy to obtain online in plain text or a more structured format such as XML.

One example project that serves thousands of books in this vein is Project Gutenberg [25]. A search of English books returns 50+ thousand results. Another mention would go to Internet Archive [26], which has a collection of millions of documents. While Project Gutenberg focuses on older world literature with expired copyright, Internet Archive serves more recent works as well. We will utilize both of these resources in our study.

The books in Project Gutenberg are highly structured in format (e.g. 'EPUB'), so they will be easy to parse and utilize; on the other hand, many books on Internet Archive are PDFs that also have OCR'ed (i.e. optical character recognition) versions, not of high quality, associated with them. We will have to clean up this type of books before we can use them.

6 Research Methodology

Graphs in general are relevant to studying language [27], the signals that exist within its expressions, and relationships between people and actions. As such, we will try utilizing graphs (made up of nodes and edges) to represent literary documents. Representation of documents as graphs has generally been done with word co-occurrence graphs [28]. In these graphs, nodes represent words in a document and edges co-occurrence of words in a given window (range) of words [29]. Co-occurrence edges are undirected, and sometimes weighted by the number of times the words co-occur. If we keep the word range to just before and after a word, the constructed graph would be a word adjacency graph.

To exemplify, consider only the following 2 sentences [30]: “Let pain be gracious. Let time be patient.” The word adjacency graph for these sentences is shown in Figure 1.

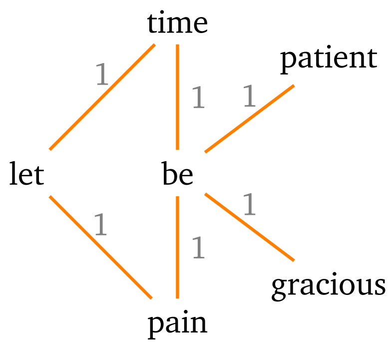

Figure 1: Undirected word adjacency graph for the two sentences “Let pain be gracious. Let time be patient.” Individual words are the nodes, without borders. Labels on the edges represent the number of times the words (i.e. incident nodes) co-occur in the sentences.

We will work on various ways in which we can represent texts as graphs that would be most relevant for bias detection. One of the methods will surely be word co-occurrence graphs, as it’s easy to implement but encodes powerful information about the individual words in sequence. This idea mirrors the famous quote by John Firth [31]: “You shall know a word by the company it keeps!” Another idea is to represent a text as a graph of syntactic dependencies, as exemplified in the Universal Dependencies project [32], where, in a sentence, each word is a node and, except the root node, depends on another word with a syntactic relation (e.g. ‘nsubj’ for nominal subject), labeled as an edge.

As there are many types of linguistic information (e.g. part-of-speech tags, such as ‘noun’) we can gather from the surface form of a text, we will try to combine the information in different ways in our experiments to learn the most relevant representations. One limitation that utilizing various linguistic information imposes on the study is that the method becomes restricted to the available tools that can extract the corresponding information. This, in turn, implies that the method won't be available to be used in most languages. For this reason, we plan to constrict the external tools to a minimum, zero if possible.

Literary texts have long narratives, with various characters and many actions taken. Narratives will be made up of sentences, in turn, made up of words. We will try to condense the information in these sentences into a graph, where nodes may represent phrases, i.e. a single word or multiple words. Edges will represent relationships between these nodes, such as the relation of one node of an adjectival phrase (e.g. beautiful) describing another node of an agent, i.e. a person or a group of people. Bias may be present in the relationships (edges) between nodes, where nodes of descriptive phrases may be biased regarding a specific group of people.

Graph neural networks (GNNs) are a family of neural networks that handle graph-structured data [33]. They learn vector representations for nodes, and optionally edges, in graphs. After converting documents to graphs (e.g. co-occurrence), GNNs can be used to learn representations for nodes that are significant for bias detection. Using GNNs, we can learn representations for the elements of the graph in an unsupervised manner, without the need for manually annotated data. Starting from a random initialization, GNNs learn representations via message passing [34]. Message passing for a single node involves *aggregating* information from its neighbors, and *updating* its representation accordingly. Message passing is an iterative process, usually stopping when the representations converge under a certain threshold. We can then use these representations in our eventual analysis. In the analysis, one approach would be to cluster the nodes based on their representations in the relevant axes (e.g. ethnicity). For example, if the axis of consideration is gender, we can take note of nodes both semantically similar to each other and close to the nodes of 'male' and 'female'.

Another useful application of graphs to represent textual information is knowledge graphs. Knowledge graphs are databases storing graph-structured knowledge, explicit in the relationships between its objects. Many knowledge graphs exist, like manually created English WordNet [35], representing semantic relationships of words, in the sense of synonyms. As they're usually crafted by experts, they're highly reliable resources. These databases would be quite helpful in merging many words that represent similar concepts, in addition to their nuances. For example, in finding gender bias, we may bundle semantically related words (e.g. mother, daughter) in one axis (e.g. gender) to make the analysis more comprehensive. We can also use them to gather words together that can be causes of bias (e.g. lowly, dirty, etc.).

A high level overview of the proposed method is shown in Figure 2. Regarding the figure, the thesis contributions will be in the ways of graph construction and the eventual analysis of the graph representations.

7 Expected Outcomes

The expected outcomes of this study are as follows: The method that will be honed over the proposed methodology during the study will be able to detect various above-mentioned social biases in literary texts. Although the eventual goal is to understand the biases that exist in any linguistic output, when the study is complete, we expect to have a better understanding of the social biases that exist in literary texts, and in turn, understand the social contexts that produce these biases.

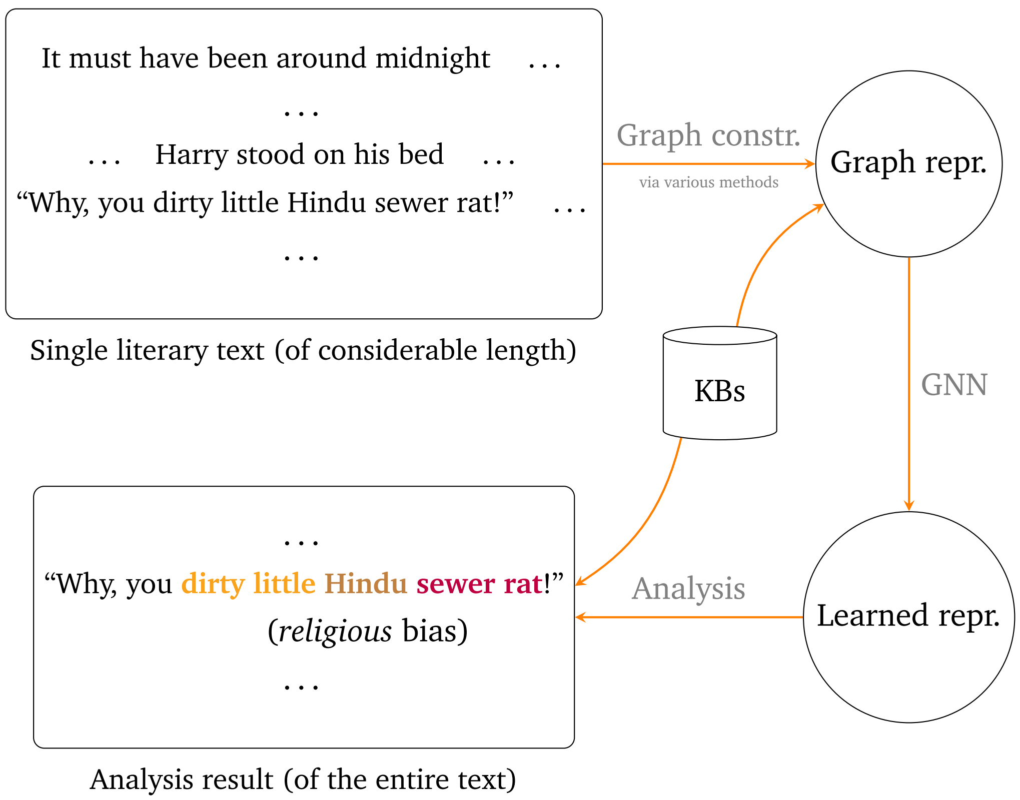

Figure 2: High level overview of the proposed method. The raw literary text [36] is, first, converted to a graph representation (e.g. co-occurrence). Then, relevant representations are learnt for the nodes via message passing (GNN). The learned representations are then used in the analysis. The phrases (**bold** above) causing the bias in the text with the corresponding types (*religious* here) are determined by the analysis. The first two words, *dirty little*, are adjectives, colored orange, and the social category of religion, *Hindu*, is the target of the bias, colored brown. The last two words, *sewer rat*, are a noun phrase, colored purple. The first and last 2 words are the descriptors of the target, causing the bias. *constr.*: construction, *repr.*: representation, *GNN*: graph neural network, *KBs*: knowledge bases.

8 Timeline

Task	Months (2024)
Literature Review *further*	Jan – Feb
Data Collection	Feb – Mar
Data Preprocessing; Knowledge Graph *start*	Mar – Apr ditto
Graph Construction	Apr – May
Graph Neural Network	May – Jul
Knowledge Graphs *again*	Jun – Jul
Bias Detection / Analysis	Jul – Sep
Thesis Writing	Sep – Nov

Table 1: Tentative timeline of the thesis

9 Conclusion

This area of study seems to be quite relevant as language has long been and is still being used by many people as a tool of power to control, restrict and use people to the speaker's benefit. Pinpointing exact sections of an output of language (either speech or written text) would be useful to people that are on the other end of the lever, the people that are spoken to. We expect to pinpoint the existence of bias in specific phrases of a given literary text. Hopefully, the output of this work will be able to further the field of bias detection and be integrated into and inspire many applications.

References

[1] Jermaine Lamarr Cole. Love yourz. Online, 2014.

[2] Francis Scott Fitzgerald. *The Great Gatsby (1925)*. na, 1991.

[3] Aesop. The fox and the grapes. Online, 2024.

[4] Harper Lee. *To Kill a Mockingbird*. J.B. Lippincott & Co., 1960.

[5] Daniel Jurafsky and James H Martin. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.

[6] Fritz Lehmann. Semantic networks. *Computers & Mathematics with Applications*, 23(2-5):1–50, 1992.

[7] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? :parrot:. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.

[8] Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

[9] Bertram Gawronski, Alison Ledgerwood, and Paul W Eastwick. Implicit bias$\neq$ bias on implicit measures. *Psychological inquiry*, 33(3):139–155, 2022.

[10] Craig S Webster, Saana Taylor, Courtney Thomas, and Jennifer M Weller. Social bias, discrimination and inequity in healthcare: mechanisms, implications and recommendations. *BJA education*, 22(4):131, 2022.

[11] Charles I Glicksberg. Bias, fiction, and the negro. *Phylon (1940-1956)*, 13(2):127–135, 1952.

[12] Adam M Croom. Slurs. *Language Sciences*, 33(3):343–358, 2011.

[13] John Steinbeck. *Of Mice and Men*. Covici Friede Publishers, 1937.

[14] Margaret Mitchell. *Gone With The Wind*. Prabhat Prakashan, 1936.

[15] Mulk Raj Anand. *Untouchable*. Penguin UK, 2014.

[16] Kimberle Crenshaw. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989:Article 8, 1989.

[17] Zehra Melce Hüsünbeyi, Didar Akar, and Arzucan Özgür. Identifying hate speech using neural networks and discourse analysis techniques. In *Proceedings of the first workshop on language technology and resources for a fair, inclusive, and safe society within the 13th language resources and evaluation conference*, pages 32–41, 2022.

[18] Hannah Kirk, Wenjie Yin, Bertie Vidgen, and Paul Röttger. SemEval-2023 task 10: Explainable detection of online sexism. In Atul Kr. Ojha, A. Seza Doğruöz, Giovanni Da San Martino, Harish Tayyar Madabushi, Ritesh Kumar, and Elisa Sartori, editors, *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2193–2210, Toronto, Canada, July 2023. Association for Computational Linguistics.

[19] Ethan Fast, Tina Vachovsky, and Michael Bernstein. Shirtless and dangerous: Quantifying linguistic signals of gender bias in an online fiction writing community. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10, pages 112–120, 2016.

[20] Chenghua Lin, Yulan He, and Richard Everson. Sentence subjectivity detection with weakly-supervised learning. In *Proceedings of 5th*

international joint conference on natural language processing, pages 1153–1161, 2011.

[21] Simon Meier-Vieracker. Racist discourse in a german far-right blog: A corpus-driven approach using word embeddings. *Discourse & Society*, page 09579265231204510, 2023.

[22] Amikul Kalra and Arkaitz Zubiaga. Sexism identification in tweets and gabs using deep neural networks. *arXiv preprint arXiv:2111.03612*, 2021.

[23] Paulina Toro Isaza, Guangxuan Xu, Akintoye Oloko, Yufang Hou, Nanyun Peng, and Dakuo Wang. Are fairy tales fair? analyzing gender bias in temporal narrative event chains of children's fairy tales. *arXiv preprint arXiv:2305.16641*, 2023.

[24] Copyright Law of the United States of America and Related Laws Contained in Title 17 of the United States Code. `https://www.copyright.gov/title17/`, 2022. Accessed: [Jan 17 2024].

[25] Project Gutenberg. `https://www.gutenberg.org/`, 2024. Accessed: [Jan 17 2024].

[26] Internet Archive. `https://archive.org/`, 2024. Accessed: [Jan 17 2024].

[27] Muskan Garg, Amit Kumar Gupta, and Rajesh Prasad. *Graph Learning and Network Science for Natural Language Processing*. CRC Press, 2022.

[28] Wenting Li, Jiahong Xue, Xi Zhang, Huacan Chen, Zeyu Chen, Feijuan Huang, and Yuanzhe Cai. Word-graph2vec: An efficient word embedding approach on word co-occurrence graph using random walk technique. In *International Conference on Web Information Systems Engineering*, pages 875–885. Springer, 2023.

[29] Ehsan Tavan and Maryam Najafi. Marsan at semeval-2023 task 10: Can adversarial training with help of a graph convolutional network detect explainable sexism? In *Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1011–1020, 2023.

[30] Adele Laurie Blue Adkins. Hold on. Online, 2021.

[31] John Firth. A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*, pages 10–32, 1957.

[32] Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. Universal Dependencies v2: An evergrowing multilingual treebank collection. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France, May 2020. European Language Resources Association.

[33] Lingfei Wu, Peng Cui, Jian Pei, Liang Zhao, and Xiaojie Guo. Graph neural networks: foundation, frontiers and applications. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4840–4841, 2022.

[34] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[35] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.

[36] Roald Dahl. *Poison (A Roald Dahl Short Story)*. Penguin UK, 2012.