Gender Bias in Fairy Tales

Lisa van Gelderen
lisa.van.gelderen@student.uva.nl
Alexia Muresan
alexia.muresan@student.uva.nl
Zoë Prins
zoë.print@student.uva.nl
Victor van der Sman
victor.van.der.sman@student.uva.nl

Abstract

Fairy tales can give children important lessons about life and the world that they live in and are mostly viewed as such. In the western world one is unlikely to find a single person that does not know the story of Snow White and many of the other famous fairy tales. These stories however often contain messages about gender roles, even though these messages might be based on outdated stereotypes (Lieberman, 1972). A lot of children grow up hearing fairy tales before bed, or they learn reading by virtue of fairy tales. As such, these stories can have a big impact on the forming of opinions in young children (Wardetzky, 1990). We map and visualize the gender bias found in fairy tales by analyzing a corpus of European fairy tales written largely by the brothers Grimm (Aho and Ullman, 1972). Using bi-grams, n-grams, a ppmi space and a Gensim Word2Vec model to show that gender bias exists on multiple levels in most fairy tales. Not only are women represented mostly as passive agents that require male characters to drive them forward, we show that female characters are largely characterized by their outward appearance instead of personality.

1 Introduction

Many of the fairy tales found today can be traced back to the bundles released by the brothers Grimm more than a hundred years ago. These stories were originally an oral tradition by female spinners in predominantly German lands. These oral stories often contained themes of rape or male oppression, and were a way for the women to stay awake while spinning. But after the brothers Grimm kept changing the stories to appeal to a larger audience, these themes were largely replaced and removed until the stories more reflected the views of 19th century patriarchal men (Marshall, 2004)

Studies about the views of children suggest their views on gender can be heavily influenced by reading certain fairy tales (Wardetzky, 1990). It is therefore worth to investigate the way these stories handle gender roles. In old stories, gender roles can be quite stereotypical and patronizing towards women. The most classic fairy tale template is a great example of this: "A brave prince needs to rescue a beautiful but helpless princess from her perils.". Gender bias in fairy tales include more focus on beauty in female characters and less on intelligence, portraying female characters as passive agents. If we teach children values with stories that contain stereotypical gender roles, these roles will be perpetuated in the next generation. This is why we aim to map the way fairy tales deal with gender in order to better understand and shed light on the problem.

2 Related Work

There has been significant research about the stereotypical portrayal of female characters in fairy tales, enforcing traditional gender roles. Across this genre women are clearly represented as the inferior gender, ingraining these biases in the minds of children who are to become yet another generation that accepts and cultivates gender bias. Central to these stories is the appearance of female characters, considered "a girl's most valuable asset, perhaps her only valuable asset" (Lieberman, 1972). Their "beauty is their wealth" (Weingart and Jorgensen, 2013), observable through lengthy descriptions of female characters' appearance, building towards a shallow representation of women. Interestingly, throughout fairy tales appears the "stereotype that a woman's character is directly tied to her appear-

ance" (Neikirk, 2013). Beauty appears to ultimately determine the value of a woman; "Girls win the prize if they are the fairest of them all; boys win if they are bold, active, and lucky" (Lieberman, 1972). This pattern repeats itself in many stories, opposing Snow White and the 'hideous' witch, Cinderella and her ugly stepsisters... Secondly, female characters' life revolves around the man in the story. Lieberman says that most of these "stories focus upon courtship, which is magnified into the most important and exciting part of a girl's life [and] marriage is the fulcrum and major event of nearly every fairy tale". It is always the happy ending, as if marriage is all a woman could ever want and need in order to be happy and fulfilled. "The passive heroine has no ambitions beyond marriage and through the lens of romance, ideas about gender roles are formed" (Meland, 2020). Most problematic is that women are depicted as a passive gender: "heroines are chosen for their beauty [...], not for anything they do [...], they seem to exist passively until they are seen by the hero" (Lieberman, 1972). This global image of women in fairy tales puta forward "traits that are supposedly inherently feminine: beauty, goodness, passivity, dependence, and an affinity for the domestic sphere" (Weingart and Jorgensen, 2013). There are few examples (if any) of strong, determined and capable female characters in fairy tales; "What is praiseworthy in males, however, is rejected in females [and] being powerful is mainly associated with being unwomanly" (Lieberman, 1972). This cultivates gender bias (criticism of "the active, ambitious, "pushy" woman" (Lieberman, 1972) that is still present nowadays) and leaves young girls without role models. Some go as far as to say that they "perpetuate Christian, patriarchal concepts as a means of maintaining the gender hierarchy" (Neikirk, 2013). Overall, fairy tales cultivate traditional gender roles, visible in the reactions of schoolchildren and teachers to an adapted production of the Princess and the Pea. Gender roles were switched, creating a princess characterised by her "physical strength and activity, her loudness, and facial expressions and lack of silence" (Meland, 2020), deemed confusing by the audience. Conversely, "the king was wearing pink and red, it was weird!' (Meland, 2020), claimed one of the teachers. This shows just how much traditional fairy tales influence what is perceived as 'normal' for both genders, "suggesting

to [children] the limitations that are imposed by sex upon a person's chances of success in various endeavors" (Lieberman, 1972).

3 Data set

The dataset used here is a fairy tale corpus from INESC-ID's Human Language Technology Lab. It is sematically organised and PoS-tagged, and contains fairy tales by five different authors, organized in clusters. The database also provides the following information:

Number of stories	453
Number of words	908,174
Average words/story	1891
Words/shortest story	75
Words/longest story	17,694
Clusters	365

In addition to the above-mentioned dataset, another dataset that we assembled ourselves with stories has also been used. This can be found in the andersongrimm.txt file. The stories in that file come from project Gutenberg.

4 Method

In order to work with the data, some additional preprocessing was performed. This consisted of putting all stories from the dataset in one list, removing non-lemmatized words, stopwords, white space and unnecessary tokens such as punctuation. Numeric characters and irrelevant word categories (conjunctions, determinants and prepositions) were filtered out, and the remaining tokens were converted to lowercase.

With the data now processed as needed, a small analysis using an n-gram model was used to generate small pieces of text based on the dataset. This was done to see if the generated text would have clear biases, which would indicate that the test the model was trained on has biases (or not) as well. Additionally, an inspection of frequent bigrams was performed to see if specific words were more commonly paired with a specific gender.

Secondly, an analysis using word embeddings was performed. This method focused on analysing surface co-occcurence by counting how many times a contextual word occurs in a collocational span surrounding the target word, in a symmetric

5 word span. After which, a similarity matrix was created based on the cosine similarity.

Finally, a Word2Vec model was trained based on the dataset to look again at the similarities between two words. Based on this model, a scatterplot was created to see how close two words are.

4.1 models

The first model used was an n-gram model trained on the Anderson and Grimm fairy tales. The models was designed to generate text of a specified length. In order to do this, the model took n-grams as input (by testing it was found that 8-grams produced the best results), and created a dictionary with the n-grams and a counter. Based on this counter, a probability was assigned to each word that that word followed another. When asked to generate text, the model choose randomly from a list of options or words to follow the previous, but gave the more likely words a higher chance of being chosen.

The second model was the Word2Vec model with parameters as specified in the Word2Vec_FastText_Comparison, and preprocessed sentences as input. For the full code, see the github repository with a single code file containing a partial narrative of all coding of the entire research: `https://github.com/lisavangelderen/FairyTale`

5 Results

As to not subject the paper to endless tables most of the results are kept in the notebook of the repository. We explain the interesting parts of our results here. In the notebook there are a number of figures that were also omitted from the paper.

In our **bi-gram analysis** we found women to be more likely to be described as beautiful than men, but less likely to be called strong or fighting. We also found the fact that women are twice as likely to be described as old, something Weingart and Jorgensen go into more depth about. Interesting about this is that while both men and women get called old as the most used adjective, women are more often described as old than they say anything. Which is telling about the descriptive role women take in fairy tales. Rather than speak or act, these women are described. This is highlighted even more in the "king" "queen" relationship. According to the co-occurrences (1 a queen can be seen as occurring mainly either next to a king or in her role as mother, with the most occurring adjective for a queen being "little" and "young". Whereas the king is "great" and has mainly active verbs in his occurrences.

	king		queen
say	7044	king	1227
son	5108	say	803
daughter	4938	gulnare	632
king	3382	go	479
come	2484	child	410
go	2344	young	397
would	1658	mother	397
great	1580	one	376
take	1557	little	339
go	1503	queen	338

Table 1: Comparing king and queen co-occurrences

In addition to bi-grams we also made use of an **n-gram model** to create short stories. Even though most of the stories were quite incoherent and could not see further than a single sentence, at some points these stories exposed bias quite clearly. One of the stories starts of with *"At home the beautiful wife prayed for her husband and her father"*, Showing how the woman cannot act but instead must hope for the return of her family (but at least she's beautiful). That same story ends when the neighbours enters through the door, again it is the male that takes action to drive the plot, and exclaims something. The final sentence being: *"To travel! to travel!. exclaimed he, overcome by most painful and passionate remembrances. But he did not despair. This was capital."*.

We also noted something interesting when looking at the **ppmi space** for the different genders. While this metric corrects for high frequency words, it showed that "woman" is mostly occurring next to "man" or "king". The same can be seen for "girl" or "queen", all of these words have a very high likelihood to be written next to their male counterparts. While this on it's own is not unusual, men and women can be used quite often together in text. The bias however comes in when reversing this analysis. For "men", "woman" is only the 7th most occurring word, for "king", "woman" is not even in the 20 most likely in the ppmi space. This trend continues with dif-
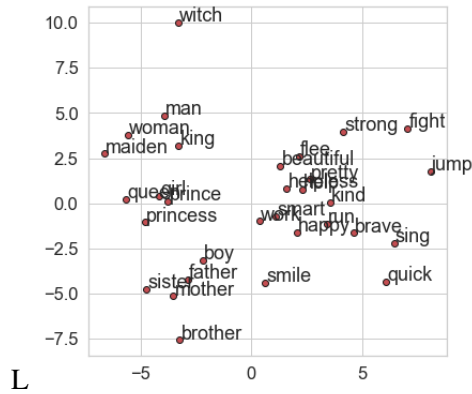
Figure 1: PCA Scatterplot created by the gensim model depicting a 2-dimensional representation of the word embeddings.

ferent male and female words such as "boy/girl", "prince/princess" or even just "male/female".

For the **Gensim model** we noted more of the same results, but to a much lesser extent than on the other metrics. With the similarities created by the model we can also see that girls are more similar to woodcutters than boys are, something we did not entirely expect as it is traditionally a job for men involving a lot of physical strength. This can also be caused by the removal of stopwords, causing these two words to be close together in the filtered data, but not in the original sentence.

Based on the wordvectors, figure 1 was obtained after a PCA.

6   Conclusion

In our paper we aim to answer the question of whether or not gender bias exists in fairy tales. This is a question with a binary response, and our response to this question is a resounding "yes". Not the least of which due to the fact that this question has been answered numerous times by other qualified researchers (Baker-Sperry, 2003) (Lieberman, 1972) (Weingart and Jorgensen, 2013). This does not devalue or paper however, as there is value in the repetition of scientific research. We hope to add something to the theory by building on this research and confirming it. In our results we found varying degrees of the same results as these previous papers.

From the bi-gram analysis we showed the dependency a queen has on her king while the king goes where he pleases. One can imagine that many young women will look up to the position of queen, as it is quite well known to be the highest

female "rank" in the story. From this implicit message these children will understand that to achieve the highest "rank" in life they will have to behave accordingly and thus stay passive. For young boys this message is quite different and has a completely different effect.

In the 2 sentences shown from the n-gram analysis the story shows all that is we find troubling with the gender bias of these models. Having learned from the fairy tales that female characters wait and look "beautiful", whereas male characters act, travel and overcome. From this we can conclude a quite clear theme in the stories where female character development is almost entirely driven by interaction with their male counterparts. While the male characters have a lot more freedom to develop and take action on their own. This is also supported by the ppmi analyses showing female words being much more inclined to be with males than vice versa.

These points add to the mounting pile of research collaborating to the fact that there is definitely certain gender bias in these stories. We found however that the results were not as obvious as previously expected, our Gensim Word2Vec model on some occasions even went completely against the grain of traditional stereotypes.

7   Future Work

Here we can give ideas on how our work can be expended on by researchers in the future that can use the information we found to do the things we did not have the resources or time to look into. There were multiple things that we found while working that, given time, we wanted to look into and explore more.

One of these things is the male gender bias in these stories. In our paper we mostly look at female oppression and stereotypes, we were not the first to do this and will likely not be the last. But the stories we looked into were mostly written/adapted in a society that wasn't entirely good for men either. Even though the men are represented and portrayed a lot more favourable in the stories than women are. The male characters are taught to be brave, this implicitly sends the message that fear is then inherently something to avoid. This, combined with more, similar, themes, add to the toxic masculinity culture present in our society today.

Another interesting idea to expand on in the fu-

ture can be the creation of stories similar to how previous work has done with children (Wardetzky, 1990). While we created short sentences using an n-gram model, this could be expanded on. Making a model, trained on fairy tales, write short fairy tale when given a certain amount of words could expose the same bias found in children.

For all results and code (not) displayed in the paper see[1].

References

Vaz Lobo, Paula and de Matos, David M. 2010. *Fairy Tale Corpus Organization Using Latent Semantic Mapping and an Item-to-item Top-n Recommendation Algorithm*. European Language Resources Association (ELRA), Malta.

Weingart, Scott and Jorgensen, Jeana 2013. *Computational Analysis of the Body in European Fairy Tales*. Literary and Linguistic Computing/ (2013): 404-416.

Aud T, Meland 2020 *Challenging gender stereotypes through a transformation of a fairy tale Challenging gender stereotypes through a transformation of a fairy tale.*. European Early Childhood Education Research Journal. 28.

Marshall E. 2004d. *The Daughter's Disenchantment: Incest as Pedagogy in Fairy Tales and Kathryn Harrison's "The Kiss."*. College English, 66(4), 403–426.

M.R. Lieberman 1972. *""Some Day My Prince Will Come": Female Acculturation through the Fairy Tale."*. College English. vol. 34, no.3, p.383

Wardetzky, K. 1990. *The Structure and Interpretation of Fairy Tales Composed by Children. The Journal of American Folklore*. The Journal of American Folklore, 103(408), 157–176.

Baker-Sperry, Lori and Liz Grauerholz 2003. *The Pervasiveness and Persistence of the Feminine Beauty Ideal in Children's Fairy Tales*. Gender and Society, pp. 711-726

Neikirk, A. 2013. *""...Happily Ever After" (or What Fairy Tales Teach Girls About Being Women"*. Anthropology, 38-42.

---

[1]https://github.com/lisavangelderen/FairyTale