

Discursive Boundaries: Characterization of Male and Female Characters in Dutch Novels, 1960s vs 2010s.¹

Theory and research question

The history of literary theory is full of grand claims about the transformative potential of literary fiction. While Plato warned us for the deceptive nature of fiction, famously arguing for banning writers from society (Republic, book X), most theorists after him stress the beneficial effects of reading literary fiction. In modern times, such theories often center on the literary genre of the novel. Lynn Hunt (2007), for instance, argues that epistolary novels such as Jean-Jacques Rousseau's *Julie* (1761) paved the way for the emergence of human rights. Arguably, readers of novels become more empathic because they wilfully put themselves in someone else's shoes – a rationale that is echoed by e.g. Nussbaum (2010), Watts (2011), and Heijne (2008). Novels, it is believed, make the world a better place.

This paper contributes to these theoretical discussions by empirically assessing the progressive potential of novels in one particular period in Dutch literary history. As a case study, it explores the role of the novel in the history of women's emancipation in the Netherlands between the 1960s and the 2010s. In light of these theoretical claims about the novel as a catalyst of social progress (Hunt 2007, Nussbaum 2010, Watts 2001, Heijne 2008), it is reasonable to expect that gender representation in novels might have been part of the changing societal position of women. In this period of approximately 50 years, which has witnessed both second and third-wave feminism, it seems safe to say that women in the Dutch language area have increasingly been better off in terms of e.g. paid labor and reproductive rights. Despite the long tradition of theorists ascribing progressive potential to novels (see above), it is, however, far from obvious how novels have concretely impacted the history of women's emancipation from the 1960s onwards.

How does the representation of gender transform in Dutch novels between the 1960s and 2010s? While being critical of the idea that social progress necessarily evolves linearly, this paper hypothesizes that novelistic representations of gender do *not* significantly change in parallel with the improved position of women in society. This expectation is based on previous comparative analyses on these two corpora, which demonstrated that female characters do not become more important in terms of network centrality in this period. (Smeets 2022). Whereas this previous study focused on network centrality as a proxy for gender representation, the present paper centers on *characterization*, which is inspired by a study by Ted Underwood et al (2018) on changes in gender representation in English language literature. If novels have played a crucial part in second- and third-wave Dutch feminism (as is suggested by e.g. Meijer 1999), this may be visible in the (changing) ways in which female characters are described as opposed to male characters. Are authors

¹ All software and metadata will be made available on my GitHub page.

increasingly using different words to characterize men and women? What do these potential differences tell us about the novel's position in the Dutch history of women's emancipation?

Data and method

In order to test this hypothesis, I have created a NLP-pipeline to trace and compare characterization of male and female characters in Dutch language novels in the 2010s and the 1960s, using two highly representative corpora: Libris2013 corpus (Smeets 2021, Smeets 2022)² and the Sanders 1960s corpus (Smeets 2022).³ Libris2013 consists of all 170 submissions to one year of the Libris Literatuurprijs, the most prestigious literary prize in the Dutch language area, and represents 37% of all originally published novels in the Dutch language in that particular year. Sanders 1960s consists of 155 novels from the early 1960s, which are part of the private digitized archive of Ewoud Sanders,⁴ and which represents 27% of Dutch-language literary fiction originally published in five years in the early 1960s (1961, 1962, 1963, 1964, 1965).⁵

The pipeline consists of five phases:

1. Automatic character detection
Working with software created for previous research (Smeets 2021, Smeets 2022), characters are automatically detected based on occurrences of their aliases throughout the texts. For the aliases, silver standard metadata is used based on manual annotations in combination with automatic enrichment (Smeets 2021, 2022)⁶. Then, Ucto⁷ is used to split the texts in the corpora into sentences. Finally, aliases of characters are replaced by their unique identifiers.
2. Ascribing Gender to Characters
Using the same silver standard metadata (Smeets 2021, Smeets 2022), gender is automatically ascribed to detected characters.⁸
3. Computing 'character text'
In order to measure characterizations of male and female characters in all novels from the two corpora, the unique identifier of each character is traced

² For metadata on this corpus, see:

https://github.com/roelsmeets/character-networks/blob/master/BOOKS_complete.csv

³ For metadata on this corpus, see: https://github.com/roelsmeets/actual-fictions/blob/main/BOOKS_AF.csv

⁴ Many thanks to Ewoud for giving me access to his archive.

⁵ It should be mentioned that, like Western literary production more broadly, Dutch literary production grew extremely between the 1960s and the 2010s. This is exemplified by the percentages mentioned here. While the samples are relatively similar in absolute numbers (155 versus 170 books), the Libris 2013 corpus covers a larger percentage of a shorter period than the Sanders 1960s corpus (one versus five years of literary production).

⁶ For the 2010s corpus, see:

https://github.com/roelsmeets/character-networks/blob/master/NAMES_complete.csv. For the 1960s corpus, see: https://github.com/roelsmeets/actual-fictions/blob/main/NAMES_AF.csv.

⁷ <https://languagemachines.github.io/ucto/>

⁸ For the 2010s corpus, see:

https://github.com/roelsmeets/character-networks/blob/master/NODES_complete.csv. For the 1960s corpus, see: https://github.com/roelsmeets/actual-fictions/blob/main/NODES_AF.csv

throughout the text. All the words in the ‘windows’ (defined as a range of two sentences) in which these identifiers occur, are appended to collections of respectively male and female characterizations dependent on the gender of the detected character. Finally, all these characterizations are aggregated on the corpus level and divided into male and female word collections.

4. Tokenizing, tagging and lemmatizing

For each corpora, the two collections of male and female characterization are tokenized, tagged and lemmatized using Frog⁹ and the results are outputted in Folia-XML files.

5. Computing loglikelihood

For the four collections of male and female characterizations (two for each corpus), the Folia-XML files are converted to plain text files containing lemmatized text. In order to get an overview of the keyness of words in these lemmatized texts, the log likelihood value of every token in the text is computed.¹⁰

Using this pipeline for both corpora will result in four collections of words that can be considered a proxy for characterization of characters:

1. Male characterization in the 2010s (Libris2013 corpus)
2. Female characterization in the 2010s (Sanders 1960s corpus)
3. Male characterization in the 1960s (Libris2013 corpus)
4. Female characterization in the 1960s (Sanders 1960s corpus)

In order to analyze differences across time and between character gender, the log likelihood value of tokens are compared across each of these collections. By doing so, this paper will reflect on possible changes in characterization between the 1960s and the 2010s and the role of gender in these changes.

Interpretation and discussion

Currently, I am in the process of fine tuning the pipeline and the first analyses will be run in the course of this month. In my talk, I will address the performance of this pipeline and evaluate my approach by interpreting the results in light of my hypothesis. My hypothesis, as stated above, is conservative. Based on an earlier comparative analysis of these corpora using network centrality as a proxy for gender representation (Smeets 2022), I expect that gender representation in novels do not evolve in tandem with the societal changes of second- and third-wave feminism. More concretely, I expect that discursive boundaries between men and women keep being visible on the level of characterization in this period of Dutch literary history. In terms of keyness, this would mean that the difference in log likelihood values of the, say, top 1000 tokens between male and female characterization in the 1960s would be relatively similar to this difference in the 2010s. Conversely, if the difference becomes

⁹ <https://languagemachines.github.io/frog/>

¹⁰ For this purpose, I have used this code: <https://github.com/mikesuhan/keyness/tree/master/keyness>

smaller overtime, that may be an indication that the cultural category of gender becomes increasingly less delineated and more fluid in novels.

Special attention will be paid to my operationalization of ‘characterization’ and how it compares to similar approaches, such as those by Underwood et al 2018. Furthermore, I will go into the merits and pitfalls of using log likelihood as a metric to assess differences and similarities in characterization.

References

Heijne, B. (2008). *Echt zien: Literatuur in het mediatijdperk*. Athenaeum - Polak & Van Genneep.

Hunt, L. (2007). *Inventing human rights: A history*. WW Norton & Company.

Meijer, M. J. (1999). *Machtige melodieën: populaire teksten uit de jaren vijftig en zestig als bron voor cultuurgeschiedenis*. Universiteit Maastricht.

Nussbaum, M. C. (2010). *Not for profit*. Princeton university press.

Plato (1957). *Republic. Book X*. Bradda Books.

Rousseau, J. J. (2013). *Julie ou la nouvelle Héloïse*. Culture commune.

Smeets, R. (2021). *Character Constellations: Representations of Social Groups in Present-Day Dutch Literary Fiction*. Leuven University Press.

Smeets, R. (2022). *Actual Fictions: Literary Representation and Character Network Analysis*. Cambridge University Press.

Underwood, T., Bamman, D., & Lee, S. (2018). The transformation of gender in English-language fiction. *Journal of Cultural Analytics*, 3(2).

Watt, I. (2001). *The rise of the novel*. Univ of California Press.