Putting Dutchcoref to the Test: Character Detection and Gender Dynamics in Contemporary Dutch Novels

Joris van Zundert[1], Andreas van Cranenburgh[2] and Roel Smeets[3]

[1]Department of Computational Literary Studies, Huygens Institute, The Netherlands
[2]Department of Language Technology, University of Groningen, The Netherlands
[3]Department of Modern Languages and Cultures, Radboud University, The Netherlands

Abstract

Although coreference resolution is a necessary step for a wide range of automated narratological analyses, most of the systems performing this task leave much to be desired in terms of either accuracy or their practical application in literary studies. While there are coreference resolution systems that demonstrate good performance on annotated fragments of novels, evaluations typically do not consider performance on the full texts of novels. In order to optimize its output for concrete use in Dutch literary studies, we are in the process of evaluating and finetuning Dutchcoref. Dutchcoref is an implementation of the Stanford Multi-Pass Sieve Coreference System for Dutch. Using a "silver standard" of annotated data on 2,137 characters in 170 contemporary Dutch novels, we assess the extent to which Dutchcoref is able to identify the most prominent characters and their gender. Furthermore, we explore the usability of the system by exploring a specific narratological question about the gender distribution of the characters. We find that Dutchcoref is highly accurate in detecting noun phrases, proper names, and pronouns referring to characters, and that it is accurate in establishing their gender. However, the ability to cluster co-references together in a character profile, which we compare to BookNLP's performance in this respect, is still sub-optimal and deteriorates with text length. We show that, notwithstanding current state of development, Dutchcoref can be applied for meaningful literary analysis, and we outline future prospects.

Keywords

character detection, coreference resolution, gender resolution Dutch literature, narratology

1. Introduction

Characters are one of the primary building blocks of narratives: their subjectivity, complexity, and agency are what sets stories in motion. For that reason, the concept of 'character' — "a text- or media-based figure in a storyworld, usually human or human-like" [9] — is among the most fundamental in narratology, next to e.g., plot, discourse, narration, focalization and motifs. To a greater or lesser extent, understanding narratives is thus understanding characters.

CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France
joris.van.zundert@huygens.knaw.nl (J. v. Zundert); a.w.van.cranenburgh@rug.nl (A. v. Cranenburgh); roel.smeets@ru.nl (R. Smeets)
0000-0003-3862-7602 (J. v. Zundert); 0000-0002-4545-1548 (A. v. Cranenburgh); 0000-0003-1656-8816 (R. Smeets)

The importance of this unit-of-analysis is exemplified by a wide range of narratological studies that rely heavily on a particular understanding of characters (cf. for instance [5]).

In order to arrive at a broader, empirical understanding of the concept, scholars have tried to automate the analysis of characters, often by doing some form of automatic character detection in larger-scale corpora (e.g. [17]). Where and how characters occur in texts is key for any sort of computational narratology; it underpins a wide range of automated analyses such as character network analysis (e.g. [15, 14]), sentiment analysis (e.g. [7]), and characterization (e.g. [18, 16]).

Character detection relies on (a variant) of coreference resolution. References and identity linking in general are one of the major challenges in artificial intelligence (AI). The goal of coreference resolution is to distinguish linguistic entities by disambiguating all individual references to them. In Tolkien's *The Lord of the Rings* one would for instance want to have indicated that the references "Frodo" and "The ring-bearer" point to the same character, and that specific instances of "Mr. Baggins", and "he" do too. Some coreference resolution systems such as, for instance, BookNLP [2] and the Stanford CorefAnnotator [13] are successful from a purely linguistic point of view. However, there is much to gain in terms of their use for specific narratological goals. Literary scholars often need more specific information than a list of the coreferences of all linguistic entities in a text. In order to be narratologically useful, coreference resolution systems have to account for some basic properties of characters. Roughly speaking, characters have at least the following properties:

1. one or more names or other identifiers;
2. humanness or animacy;
3. fulfilling a function in the narrative (e.g. the subject, object, sender, helper [8]);

For the widely used BookNLP an F1 of 79.0 was reported [2] on tasks 1 and 2. Thus, a system like BookNLP is able to identify a wide range of coreferences of entities in a text. However, further sophisticated steps are required to fully capture the narratological richness of characters. Especially the identification of narrative function of characters is still a rather open task. Characters are part of a hierarchically structured fictional world where some characters are more central than others [15]. This means that not every (named) entity is equally meaningful in literary texts. Not all entities identified by, for instance, BookNLP are considered characters from a narratological point of view, as they often do not fulfill a meaningful function in the storyworld. Commonly therefore, character analysis applies some form of network analysis to identify relevant characters [15, 1, 6].

BookNLP is currently only available in English, and although it has acquired funding to expand the number of supported languages, Dutch is not among the prospective supported languages. The development of a narratological informed and accurately performing BookNLP-like system for the Dutch language thus remains a strong desideratum for Dutch literary research.

Dutchcoref, an implementation of the Stanford Multi-Pass Sieve Coreference System for Dutch literature [3, 4] is a first step towards this ideal. The goal of the present paper is to evaluate the accuracy and usefulness of Dutchcoref given the tasks 1, 2, and 3. Using a "silver standard" of annotated data on 2,137 characters in 170 contemporary Dutch novels, we assess

the extent to which Dutchcoref is able to identify the most prominent characters and their gender. We also assess the narratological relevance of the system by exploring the ability to provide insight about gender distribution and dynamics of the characters in the novels of the corpus.

Earlier evaluations of Dutchcoref demonstrated that in a dataset of contemporary Dutch literature, 95.0% of human mentions[1] are correctly recognized; furthermore, 89.9% of male mentions, and 73.4% of female mentions can be distinguished [4, p.51]. Taking those results into account, in this paper we present an analysis of a larger corpus of contemporary novels, to analyze the distribution of human mentions, as well as their gender (im)balance and dynamics.

2. Data and Method

For our twofold evaluation of Dutchcoref we use the Libris2013 corpus, consisting of all 170 submissions in one year of the annually awarded Libris Literatuurprijs, one of the most prestigious literary prizes in the Dutch language area [10, p. 15]. All novels were published in 2012, were written in the Dutch language, and are considered literary novels (with the corresponding NUR[2] code 301.) These 170 books represent 37 percent of all the novels published in that particular year.

Earlier research on this corpus [5, 15, 14] has resulted in an extensive "silver standard" metadata on 2,137 semi-automatically identified characters, consisting of demographic information such as gender, age, education, cultural background, and profession. A combination of automated and manual text analysis was necessary to create a dataset that contains all of the characters that fulfill a narratological function in the dataset. Aliases of each of these 2,137 identified characters were collected on the basis of name variants only, thus excluding other coreferents (pronouns and descriptions such as 'the man who walks down the street'). Therefore we call this a "silver standard" because it is not based on full annotation of all coreferences, but represents those characters (and their named aliases) that are meaningful from a narratological point of view. The silver standard has been used for automatic character network analysis [15, 14] and text mining applications [12]. Although a modest "gold standard" containing a sample of 21 novels in Dutch exists for which all coreference has been annotated [3, p. 41], this silver standard offers a comparatively large corpus of 170 full texts from Dutch novels in which all named characters have been identified, also compared to typical English evaluation data (e.g. [6] in which for 40 novels 300 annotated sentences each were used).

In section 3, we use this silver standard to evaluate the accuracy of Dutchcoref. The silver standard allows us to identify all positions in the texts of the novels where a person is referenced by name or name variant. All novels are analyzed with Dutchcoref which results, among other analytical data, in a CoNLL file with coreference information and a file listing all "mentions", that is: all one or multiple token fragments that Dutchcoref identifies as a mention of a person, location, organization, etc. We assess to what extent Dutchcoref is able to identify all silver data

[1] A mention refers to an instance of a description, name, or pronoun referring to a person or object mentioned in the text.

[2] NUR is a marketing instrument used by publishers and booksellers to categorize books with an eye on the different sections in bookshops. According to [19, p. 400] it would appear that "NUR can be regarded as a rough approximation of the concept of genre as it is understood by booksellers and readers."

name variants. We then continue to compare its performance to that of BookNLP. In section 4 we progress to examine how well Ductcoref in its current state can be applied for a meaningful literary analysis of the Libris 2013 corpus.

3. Evaluation of Dutchcoref against a "silver standard"

Dutchcoref was used to analyze the text of all novels. Dutchcoref yields (among other analytical results) a CoNLL file providing lexical, syntactic, and coreference information for all tokens in a text. We use the CoNLL file to locate all name variants listed by the silver data. Dutchcoref also yields a file containing "mentions", listing all single and multi-token occurrences in a text of which Dutchcoref asserts they are nouns (or noun phrases), proper names, or pronouns referring to persons, locations, organisations, etc. Combined into a match table this information allows us to compute recall. Note that we cannot compute an F-score because there is no reliable way to determine exactly what a false positive is in this case: Dutchcoref generates many more mentions (among other things in the form of pronouns) that might very well be correctly identified co-references to characters, but our silver data will not tell us if they are indeed correct. Therefore, we rather compute two recall scores, a strict and a lenient score. The strict recall only scores exact matches between silver standard and Dutchcoref identified characters, while the lenient score allows for some leeway in the span of the text identified by Dutchcoref. Figure 1 provides an example of lenient and and strict matching. In the interest of precision we provide the strict measure, although a human reader would easily confirm that in most instances by far the lenient matches should indeed be counted as correct.

Position	Silver	Dutchcoref	Evaluation	
			lenient	strict
44449	Brady	Brady	match	match
64131	Brady	de heer Brady	match	no match
19957	Hamid	Hamid Shakir Mahmouds	match	no match
31550	Hamid	Die Hamid Shakir Ahmed	match	no match
5278	Hamid Shakir Mahmoud	Hamid Shakir Mahmoud	match	match
16102	Hamid Shakir Mahmoud	Hamid Shakir Mahmoud	match	match
16146	Hamid Shakir Mahmoud	een andere Hamid Shakir Mahmoud	match	no match

Figure 1: Selection of a match table showing typical strict and lenient matches between "silver data" and Dutchcoref mentions.

Across the corpus as a whole we find a strict recall of 0.90, and a lenient recall of 0.97. A density plot of recall results (figure 2) further corroborates the high accuracy of Dutchcoref in identifying character name variants.

Dutchcoref performs well as a named entity recognition (NER) tool "on steroids". It does not just yield proper names indicating characters, such as "Böckli" and "Fehmer". Rather, it also identifies descriptive phrases that refer to the same characters, such as "de heer Böckli" (en.: "Mr. Böckli") and "de beroemde architect Fehmer" (en.: "the famous architect Fehmer"), and predicts properties such as gender and number for them. Unfortunately, without a gold standard we cannot determine exactly how accurate Dutchcoref performs on this corpus in

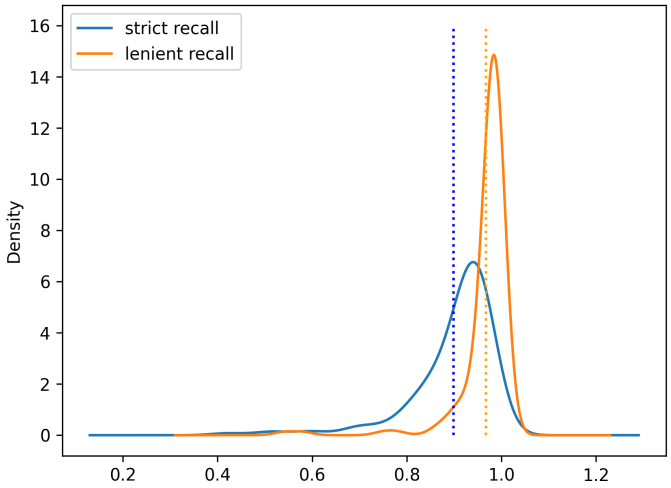

Figure 2: Density plot of recall scores on novels across the corpus.

this respect, but perusing match tables for different novels indicates that most "no matches" are actually correct identifications of compounded co-references.

Ideally a co-reference resolution tool goes beyond identifying proper names and noun phrases that refer to characters, and in addition also identifies correctly which pronouns refer to what characters. Dutchcoref tries to construct clusters of mentions that pertain to one character or object. Without a gold standard it is very hard to tell Dutchcoref's accuracy in this respect, but we can gauge its performance a little from its behavior across shorter and longer text samples. What we observe is that Dutchcoref tends to combine too many references in one cluster if the length of the analyzed text increases. Thus, in short samples of text it seems to be functioning reasonably well, yielding groups or clusters of mentions that pertain to one particular character, and ideally these would represent meaningful characters from the novel. However, when Dutchcoref sees a longer text it starts wildly conflating references to different characters. A typical example is given in figure 3, which shows the relative contribution (y-axis) from different clusters of co-references (x-axis) to the total amount of references. The changing characteristic between the right and left chart shows how Dutchcoref assigns more and more co-references to the same cluster if it shown a larger part of the text. The table in 4 shows the (first) actual references from the largest cluster from both charts. It is easily observed that while the co-reference resolution for the sample (depicted as the bottom table) might still show some coherence, the top one has aggregated far too many mutually exclusive references into one cluster, conflating several characters in the process.

We compared this progressively worsening accuracy of Dutchcoref to BookNLP's performance, which is a neural pipeline for English literary texts. For this we selected four books (from the 19th century to current) that are not in BookNLP's training corpus (LitBank [2]), i.e. "The Girl on the Train", "The Running Man", "The Grapes of Wrath", and Dorothy

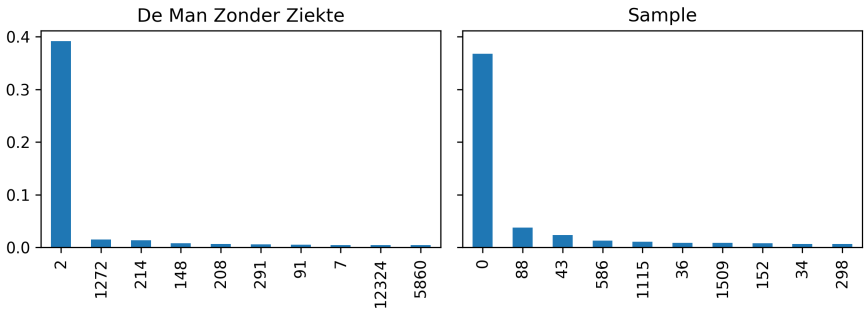

Figure 3: Contribution to total mentions across clusters (i.e., characters) in a novel from Dutch author Arnon Grunberg ("De man zonder ziekte", en. The Man Without Illness). On the left distributions for the full novel, on the right for a 10% sample of the text. The labels on the x-axis are simply IDs of identified clusters and are not meaningful in any numeric sense.

Wordsworth's "Journal Volume 1". In figure 5 we reproduce the result from Steinbeck's novel as an illustrative measure. Our results indicate that BookNLP falls victim to the same defect, but to a much lesser extent than Dutchcoref. However, we also surmise that BookNLP's greater accuracy is actually caused by it only taking into account coreferences for a restricted set of entity categories, rather than all noun phrases, as Dutchcoref does. Therefore, it might be advantageous to re-train the coreference system to consider only mentions referring to persons. Reducing the number of mentions than can be linked also reduces the potential for erroneous links.

4. Using Dutchcoref for narratological exploration

Given its current performance in clustering coreferences in long texts Dutchcoref may not be a very useful tool for accurate character identification yet. We hope to improve its accuracy over time, so that clusters will reliably coincide with characters. For the inference of social networks of characters and their interactions this is pivotal, as merely looking at named entities leaves out a lot of information: around 40% of mentions in literature are pronouns, while names are approximately 10% (with the rest being nominal descriptions) [3]. Yet, in its current state Dutchcoref already has potential to assist in narratological analysis. While its clustering mechanism clearly needs work, its ability to indicate names and pronouns is state of the art. Moreover Dutchcoref indicates whether a noun or pronoun refers to a human entity and tries to establish the entity's gender, and reaches 0.84 recall on proper names given the current silver standard. Lacking a fully annotated corpus there is little sensible evaluation possible with regard to pronouns, but in fiction most pronouns by far relate to characters, in which case gender identification is trivial in Dutch. Together this allows us substantial insight into the gender dynamics in the Dutch literary corpus we are using.

We know from prior research by, inter alia, Corina Koolen [10] that gender balance in Dutch literature is skewed heavily towards the male side of the spectrum in many respects: reader appreciation, critical appraisal, awards, etc. Koolen also looked at differences in vocabulary

De Man Zonder Ziekte		
First 90 unique mentions in all 6928 mentions from cluster 2		
Samarendra Ambani	zijn	zijn vriendin
een nieuwe koffer	Samarendra's vriendin	hij
zijn koffer op de bagageband	ze	Sam
Ze	mij	je koffer
Zijn naam	hem	zijn moeder
Zijn	Zijn vriendin	het geld
Hij	me	ik
meneer Ambani	zich	je
Sams	Sams vriendin	Ik
zijn toekomstige vrouw	Zijn moeder	een man
het kind	de man zonder ziekte	Zijn koffer
die	haar	een vrouw
Alles	zijn Zwitserse vrouw	een goede man
Zwitser onder de Zwitsers	een vriend	De vriend
Het gezin Ambani	zij	de moeder
mevrouw Ambani	de man in huis	Meneer Ambani
haar man	Ambani's	Mevrouw Ambani
Nina	Een naam	het gezin Ambani
zijn moeder in de woonkamer op de bank	Sams moeder	een dikke , stinkende man
een naam	kind	zijn studie architectuur
de vrouw	Zijn toekomstige vriendin	zijn toekomstige vriendin
een moeder	de grote , anonieme beïnvloeder van andermans geluk	dat
de architect	mijn	een geschikte vrouw voor hem
Het	de beschaafdste vrouw	uw
zijn vriendin om te zien of ze ook trots is	een iets oudere , getrouwde vrouw	de familie Ambani
een vriendin	de vrouw in de Audi en de paar studentes	zijn koffer
Nina's	De wereldberoemde architect Max Fehmer	de naam van het bureau
Fehmer	Architectuur	architectuur
De architect	hun	Max Fehmer
een architect	een toegewijd en dienend architect	Hamid Shakir Mahmoud

Sample		
Unique mentions in all 566 mentions from cluster 0		
Ik	Bill	de man met het blonde haar
Hij	zich	hij
Sam	Hassan	ik
Het	zijn	hem
je	Mijn	Een man van Erbil International
de man	De man	me
u	Zijn	mijn
Je	een man met een snor	dit
Honey	die	het
Ze	Niemand	een vrij grote badkamer
mij	mijn kleren	Sams
Dit	jouw	jouw kleren
een verkoper	Jouw	een man
Jouw kleren	iemand	andermans

Figure 4: Coreferences attributed to the two largest clusters reported in the charts of figure 3.

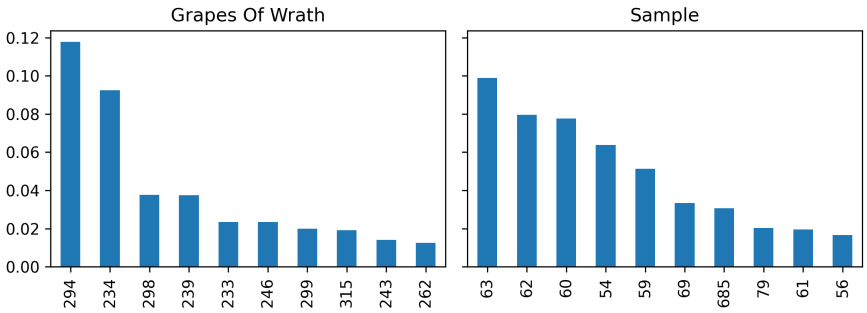

Figure 5: Distributions of mentions across clusters in a novel analyzed by BookNLP.

and topic use between male and female authors, using LIWC, topic models and purpose made extraction algorithms. However, Koolen did not consider how gender representation differs throughout the texts of novels themselves. Using Dutchcoref we can now easily add such an aspect of narratological analysis. Using a rolling window approach we can count and average the dispersion of female and male proper names and pronouns throughout full texts, creating a visual chart of gender balance across each novel. We produced such charts for windows based on 1,000 token windows (i.e. each window is roughly two pages) progressing the window through the text token by token. We also produced charts in the same manner but based on 100 paragraphs each, progressing paragraph by paragraph. These parameters were chosen to gauge if results for a method that respects text structure (i.e. paragraph boundaries) would significantly deviate from a method that does not (i.e. that is token based). Consult figures 7 through 10 for four examples. The chart in the top left of each figure shows the contribution of female names and of male names to all tokens as percentage of the 1,000 tokens in each window. The top right chart in each figure shows the same, but as the percentage of all tokens in each paragraph. The two charts at the middle level of each figure show the same measure but for male and female pronouns. The bottom charts in each figure give the use of female names (or pronouns) as the ratio of all names (or pronouns) for each window of 1,000 tokens (left) or paragraphs (right).

To gauge possible skewedness at corpus level we can aggregate the numbers for individual novels. Figure 6 shows a density graph of the difference between the use of female proper names and pronouns and the use of male proper names and pronoun in each novel. For each novel we computed the mean percentage of female and the mean percentage of male names and pronouns across all 1,000 token windows. Then we calculated the difference between these (male minus female). These give us a data point for each novel shown in the top bar of figure 6. The density plot below that is based on the values for all the novels. If genders were somewhat balanced we would expect a bell curve centered around zero on the x-axis. Instead the center is far more to the male side of the spectrum. An aggregated corpus ratio of the means confirms that male names and pronouns are used more than twice as many times as female ones. The token-based corpus-wide mean for use of male proper names and pronouns is 0.29, while for female ones it is 0.14. Numbers for the paragraph based approach are comparable and this, together with highly congruent charts, suggests that unit of measure does not play a significant

role in these results. Note that the observed clear bias does not fully conclusively show that Dutch literary production is a heavily male focused activity. As Koolen [10, p.134-6] argues, there exist known selection biases in bulk, long and short lists for literary prizes, which is the type of corpus we are working with (in this case: a bulk list of all novels submitted for a literary prize in one year). Further analysis should evaluate if the same bias exists in a random sample of contemporary literary production. The numbers we yield give us pause to think in any case.

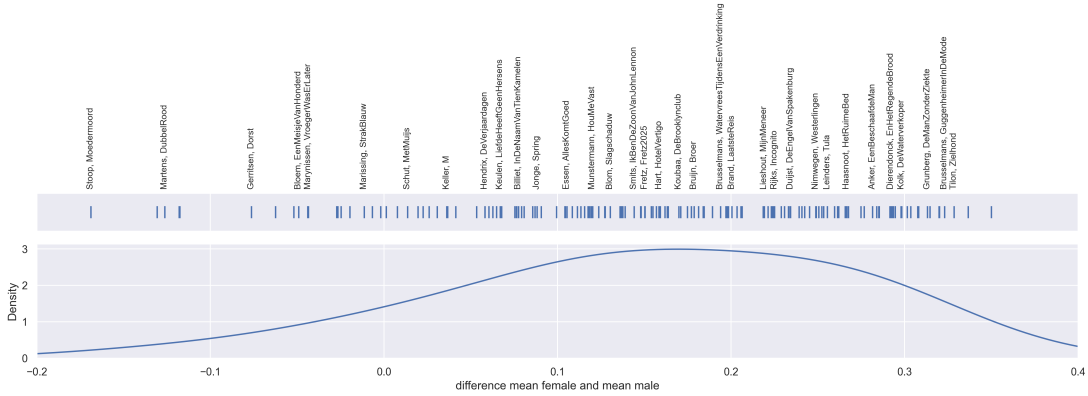

Figure 6: Density graph based on the difference between female and male proper names and pronouns use in novels. The more the apex of the curve deviates from the x-axis origin the more skewed the corpus is. Some novel labels have been dropped to avoid unreadablee labels due to overlap. We have made sure to include the labels for the novels represented in figures 7 through 10.

Qualitatively, gender dispersion plots as shown in figure 7 through 10, are useful to add a new distant reading perspective on gender representation within particular novels. The ways in which (male and female) pronouns and (male and female) proper names are distributed across a novel elicits an additional dimension that can be taken into account by studies on the literary representation of gender. Most obviously, such dispersion plots add to our understanding of gender representation in terms of visibility. Up until now, studies taking into account visibility as a factor of gender representation focus on how often respectively male and female characters occur in literary texts (e.g. [15], [11]) or analyze differences in characterization of male and female characters (e.g. [18, 12]). Narratologically, those vantage points make sense: we first need to detect where male and female characters occur in the text before we can assert any claims about their visibility. However, the rolling window measures for pronoun and proper name use shown in figures 7 through 10 are agnostic about any narratological definition of what a character is. Although we intend to retrain Dutchcoref on the 'person'-category only (in order to come closer to such narratological definitions), we can already use its output to qualitatively track the evolution of gender representation in particular works based on the rough distribution of pronouns and proper names (and potentially other noun phrases as well).

As such, these plots yield a more distributed view on the gender dynamics within particular novels than studies that focus on a more restricted definition of 'character'. More generally, density graphs for full corpora, such as figure 6 can be used to track down individual works

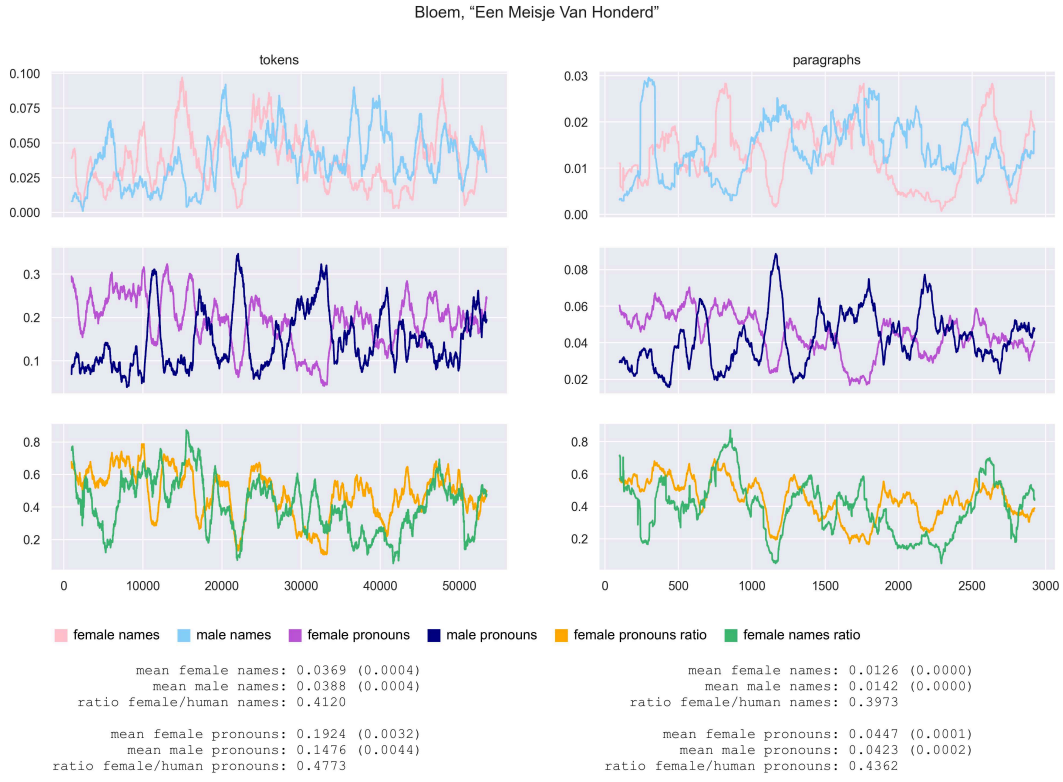

Figure 7: Proper name and pronoun dynamics in Bloem’s “Een meisje van honderd”

that either conform to or deviate from a particular norm, while rolling plots (such as figure 7) offer more insight in the particular gender dynamics of a single novel. In some novels, for instance, the trend lines do not or only barely cross (e.g. see figure 8 and 9), which suggests that the over- or under-representation of one particular gender is remarkably constant and stable throughout the narrative. For instance, Herman Brusselmans’ *Guggenheimer in de mode* (2012), occurs at the very right part of the plot in figure 6. From a close reading point of view, that makes sense. This particular novels excels at stereotypical representations of women, which has become one of the trademarks of Brusselmans’ authorship. In figure 10 a sharp distinction between male and female gender is visible, which aligns perfectly with the male-dominated, reactionary fictional world that Brusselmans has created in this particular novel. In that imaginary universe, women are constantly objectified and sexualized, and their narratological function solely seems to fulfill the goals and desires of the male protagonist. Such qualitative observations match perfectly with the dispersion of pronouns and proper names as shown in figure 10: both qualitatively and quantitatively the under-representation and stereotypical characterization of female characters seems constant. Conversely, if the the trend lines in the dispersion plots of novels do cross at particular moments in the narrative (e.g. see figure 7), that might indicate a shift in gender representation that might have various narratological

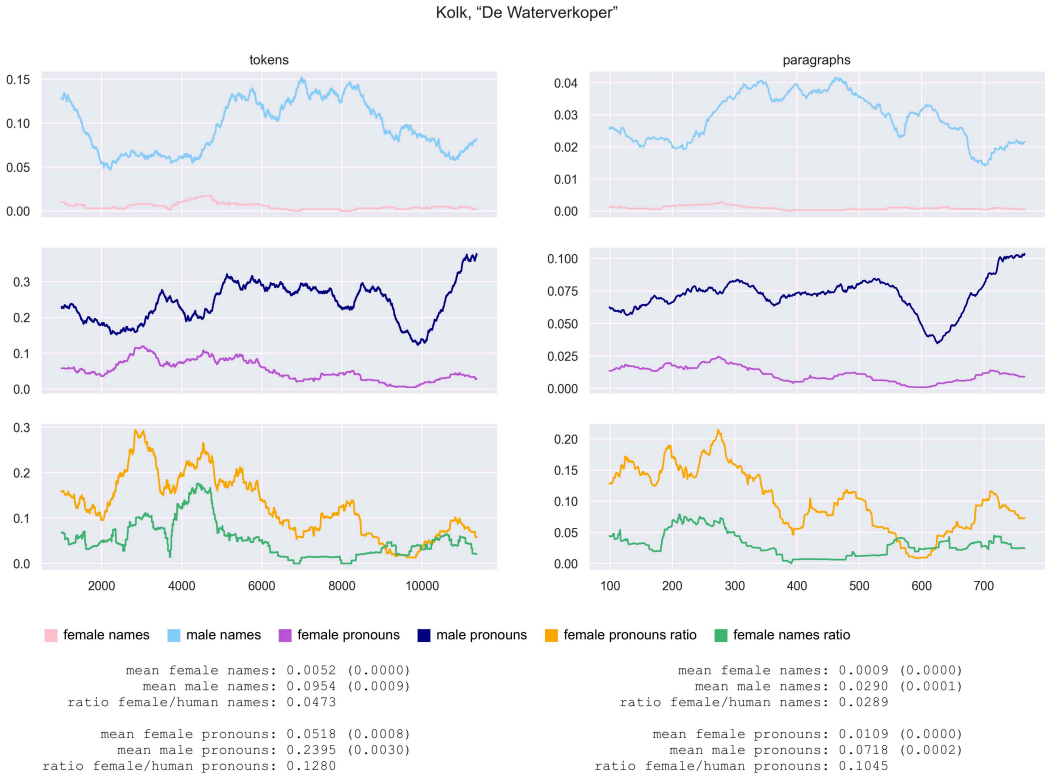

Figure 8: Proper name and pronoun dynamics in Kolk “De waterverkoper”

explanations in terms of e.g. focalization, narration, or motif structure. Figure 8, for instance, represents a gender dynamic view of Lidewij Martens’ novel *Dubbel Rood*. The parts of this graph indicating less mentions of female names seem to coincide with perspective changes in the story.

5. Discussion and Conclusion

Accurate identification of characters in narratives, including co-reference resolution of pronouns and noun phrases referring to those characters, remains a high value desideratum for computational literary research. Our contribution shows that the current state of the art in NLP for this task is insufficient to clearly, unambiguously, and accurately identify and describe characters in Dutch and English literary materials in some automated fashion. Dutchcoref shows a high recall on proper names and their gender identification, but co-reference resolution performance deteriorates quickly with longer (novel sized) text, causing too many co-references to different characters to be conflated. Although BookNLP performs better on English texts, this may be due to the fact that BookNLP ignores many types of referring expressions such

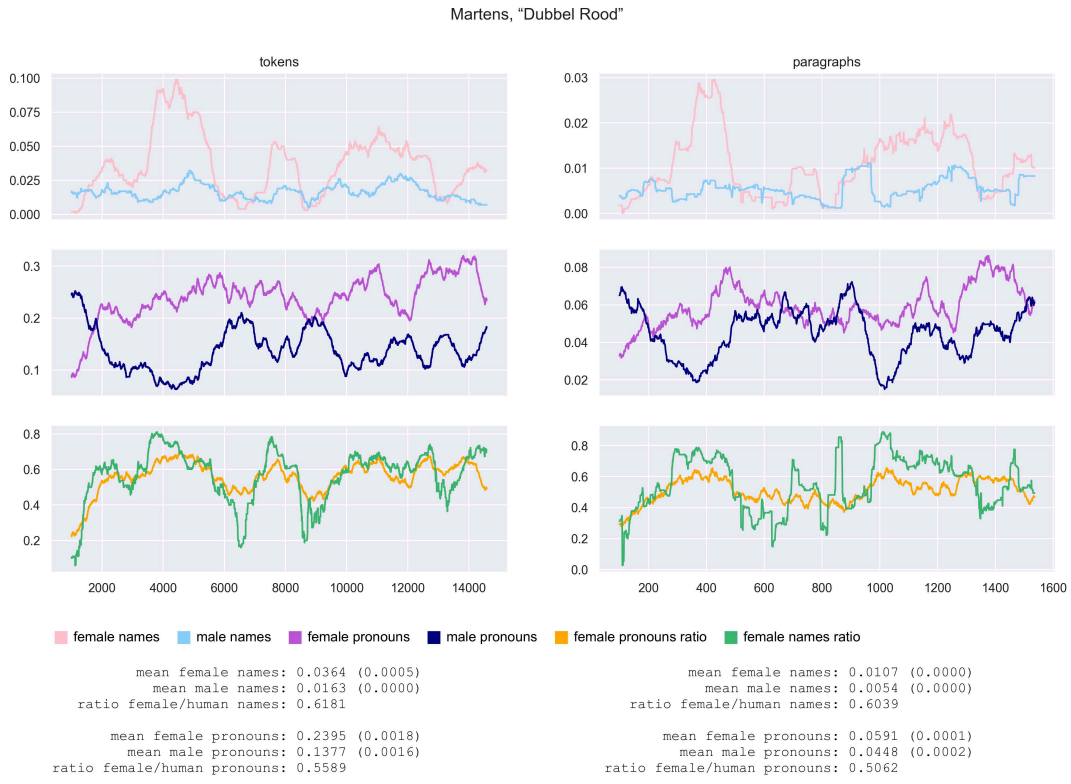

Figure 9: Proper name and pronoun dynamics in Martens “Dubbel Rood”

as noun phrases describing objects. This behavior may negatively impact BookNLP’s usefulness for narratological analysis. From the current discourse it is not clear whether such noun phrases (e.g. ‘the ring’, ‘photographs’) will be highly contributory to tasks in computational literary research, although it seems plausible that accurate identification and resolution of such noun phrases will be important for (automated) event and plot analysis.

Their current sub-optimal performance does not preclude useful application of BookNLP and Dutchcoref in the domain of (computational) literary analysis. We have shown that with its current abilities Dutchcoref can positively contribute to, for instance, the granularity of our measurement and therefore knowledge about character gender dynamics in novels. This may well impact the current state of the art in social network analysis of novel character, as well as event analysis.

We were able to add our insights by using a “silver standard” listing the proper names and gender properties of 170 Dutch contemporary novels. To improve the performance of Dutchcoref, which we consider as one of our challenges for future work, a gold standard of a fully annotated corpus – including explicit information on the relations between all referents and antecedents – remains as much a desideratum as highly accurate co-reference resolution for literary texts.

Code & Data Availability

The data of the silver standard are available in a GitHub repository: https://github.com/roelsmeets/character-networks. Data and Jupyter notebooks (Python) for our evaluation and analyses are also available via GitHub: https://github.com/jorisvanzundert/computational_characters.

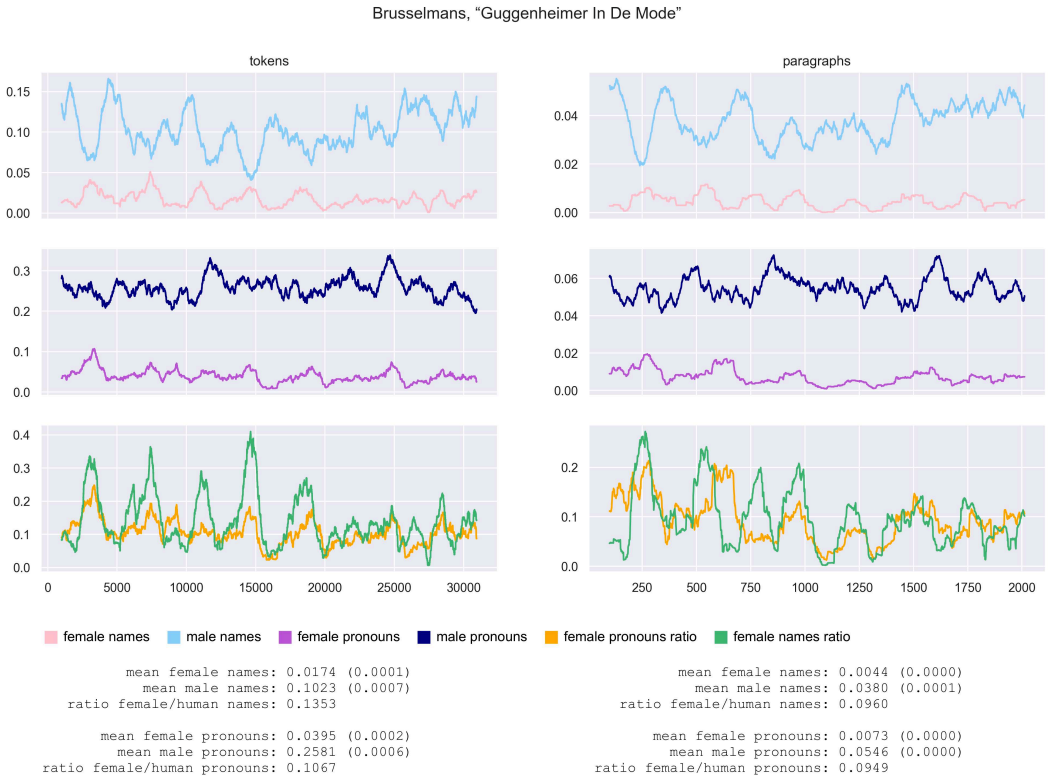

Figure 10: Proper name and pronoun dynamics in Brusselmans’ “Guggenheimer in de mode”

References

[1] A. Agarwal, A. Kotalwar, and O. Rambow. “Automatic Extraction of Social Networks from Literary Text: A Case Study on Alice in Wonderland”. In: *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Nagoya: Asian Federation of Natural Language Processing, 2013. URL: https://api.semanticscholar.org/CorpusID:6829505.

[2] D. Bamman. *BookNLP*. Codebase published on GitHub. Berkeley, 2021. URL: https://github.com/booknlp/booknlp.

[3] A. van Cranenburgh. “A Dutch coreference resolution system with an evaluation on literary fiction”. In: *Computational Linguistics in the Netherlands Journal* 9 (2019), pp. 27–54. URL: https://clinjournal.org/clinj/article/view/91.

[4] A. van Cranenburgh, E. Ploeger, F. van den Berg, and R. Thüss. “A Hybrid Rule-Based and Neural Coreference Resolution System with an Evaluation on Dutch Literature”. In: *Proceedings of CRAC*. Punta Cana, Dominican Republic, 2021, pp. 47–56. URL: https://aclanthology.org/2021.crac-1.5.

[5] L. van der Deijl, S. Pieterse, M. Prinse, and R. Smeets. “Mapping the Demographic Landscape of Characters in Recent Dutch Prose: A Quantitative Approach to Literary Representation”. In: *Journal of Dutch Literature* 7.1 (2016), pp. 20–42. URL: https://journalofdutchliterature.org/index.php/jdl/article/view/103.

[6] N. Dekker, T. Kuhn, and M. van Erp. “Evaluating named entity recognition tools for extracting social networks from novels”. In: *PeerJ Computer Science* 5 (2019). Ed. by D. Amancio, e189. URL: https://doi.org/10.7717/peerj-cs.189.

[7] E. Gius, A. Murawska, O. Schmidt, C. Sökefeld, and M. Vauth. “Sentiment sensitivity. Using sentiment analysis in literary studies to analyze genre and the depiction of illness”. In: *DH2020 correfours/intersections – Digital Humanities 2020 Book of Abstracts*. Ottawa: ADHO, University of Ottawa, and Carleton University, 2020, pp. 216–219. URL: https://bit.ly/3O0FL3i.

[8] A. J. Greimas. “Éléments pour une théorie de l’interprétation du récit mythique”. In: *Communications* 8.1 (1966), pp. 28–59.

[9] F. Jannidis. “Character”. In: *The living handbook of narratology*. Ed. by P. e. a. Hühn. Hamburg: Hamburg University, 2013. URL: http://www.lhn.uni-hamburg.de/article/character.

[10] C. Koolen. “Reading Beyond the Female: the relationship between perception of author gender and literary quality”. PhD thesis. Amsterdam: University of Amsterdam, 2018. URL: https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37f1ce7.

[11] E. Kraicer and A. Piper. “Social characters: the hierarchy of gender in contemporary English-language fiction”. In: *Journal of Cultural Analytics* 3.2 (2019). URL: http://doi.org/10.22148/16.032.

[12] J. Neugarten and R. Smeets. “MythFic Metadata: Exploring Gendered Violence in Fanfiction about Greek Mythology”. In: *Proceedings of DHBenelux*. Brussesl: Kbr, 2023. URL: https://2023.dhbenelux.org/wp-content/uploads/2023/05/DHB2023%5C%5Fpaper%5C%5F2257.pdf.

[13] K. Raghunathan, H. Lee, S. Rangarajan, N. Chambers, M. Surdeanu, D. Jurafsky, and C. Manning. “A Multi-Pass Sieve for Coreference Resolution”. In: *Proceedings of EMNLP*. Cambridge, MA, 2010, pp. 492–501. URL: https://aclanthology.org/D10-1048.

[14] R. Smeets. *Actual Fictions: Literary Representation and Character Network Analysis*. Elements in Digital Literary Studies. Cambrdige: Cambridge University Press, 2022. DOI: 10.1017/9781009180788. URL: https://doi.org/10.1017/9781009180788.

[15] R. Smeets. *Character Constellations: Representations of Social Groups in Present-Day Dutch Literary Fiction*. Leuven: Leuven University Press, 2021. URL: https://library.oapen.org/handle/20.500.12657/51177.

[16] R. Smeets. "Discursive Boundaries: Characterization of Male and Female Characters in Dutch Novels, 1960s vs 2010s." In: *Digital Humanities Benelux abstracts*. Brussels: Kbr, 2023. URL: https://2023.dhbenelux.org/wp-content/uploads/2023/05/DHB2023%5C%5Fpaper%5C%5F2127.pdf.

[17] R. Smeets. "Finding characters: An evaluation of named entity recognition tools for {Dutch} literary fiction". In: *European Alliance for Digital Humanities (EADH) Day 2017*. Rome, Italy: Eadh, 2017.

[18] T. Underwood, D. Bamman, and S. Lee. "The transformation of gender in English-language fiction". In: *Journal of Cultural Analytics* 3.2 (2018). URL: http://doi.org/10.22148/16.019.

[19] J. Van Zundert, M. Koolen, J. Neugarten, P. Boot, W. Van Hage, and O. Mussmann. "What Do We Talk About When We Talk About Topic?" In: *Proceedings of the Conference on Computational Humanities Research 2022*. Antwerpen: CEUR Workshop Proceedings, 2022, pp. 398–410. URL: https://ceur-ws.org/Vol-3290/.