Character Constellations

Smeets, Roel

Published by Leuven University Press

Smeets, Roel.


For additional information about this book
https://muse.jhu.edu/book/98048

For content related to this chapter
https://muse.jhu.edu/related_content?type=book&id=3191736
CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION: CHARACTER CONSTELLATIONS

Like the stars in the sky, the arrangement of fictional characters in literary texts form patterns we might call ‘constellations’. Constellations are products of imagination: they are not just there but emerge in the eye of the beholder. Ancient cultures have looked closely at the celestial sphere and discovered images of animals, gods, and mythological creatures in the arrangement of the stars and interpreted those images in light of specific time- and place-related norms, values, and beliefs. In a similar vein, yesterday’s, today’s, and tomorrow’s readers ascribe different meanings to the patterns they discover in the interrelations between characters in narrative fiction. Unlike the stars, however, characters populating books do not change in shape or position. Whereas constellations of stars eventually – although very slowly – transform into different images due to the stars’ relative position to planet earth, character constellations remain the same with the passing of time because characters occur in fixed positions in the pages of the book. Unless a story is rewritten, characters will not move an inch from their position in the narrative’s linguistic structure. But readers change and so do the images they discover in the relative positions characters occupy in texts.

This book is about the ‘images’ of social groups of people that can be discovered in the character constellations in present-day Dutch literary fiction. Through the metaphor of character constellations, it attempts to ground discussions on literary representation in the immovable, fixed positions of characters at the
sentence level while leaving room for different interpretations of the images of social groups these characters represent. Just like the ancient people searched for patterns in the stars to gain insight into the world around them, unraveling patterns between fictional characters contributes to a better understanding of the literary representations of people circulating in the world in which these books were written. A deeper insight into depictions of particular groups of people in recent Dutch language fiction may shed light on norms, values, and beliefs associated with the social dichotomies of, for instance, men as opposed to women, people with or without a migration background, the less as opposed to the higher educated, and the young as opposed to the old.

As the literary representation of people with different demographic backgrounds and identities has been subject to heated ideological discussions, the present study ventures into precarious waters. In academia these discussions take place within the critique of literary representation focusing on characters (see section 1.2 of this chapter). Through detailed close readings, this branch of literary criticism has disentangled various representations of people with a particular gender, descent, race, ethnicity, class, age, sexuality, or other identity category in light of Marxist, feminist, postcolonial, or other ideologically oriented theoretical strands. Often, such studies have – more or less convincingly – argued that hierarchies, biases, and inequalities are apparent in the ways in which different social groups are depicted in particular texts. What these discussions have lacked up to the present is a methodology for measuring the relative positions of characters in the narrative similar to the tools astronomers use to measure the relative positions of stars.

This book argues that such a methodology is crucial for a broader and deeper understanding of representations of social groups. One of its premises is that ‘images’ of social groups arise in the interrelations between characters, just as stars form patterns only in relation to other stars. Building on that idea, it proposes, applies, and evaluates a range of data-driven, statistical models to trace the interrelations between 2,137 identified characters in a sample corpus of 170 Dutch books of literary fiction published in 2012, for which extensive metadata (gender, descent, educational level, age, profession) were gathered (see section 1.4.2 of this chapter for more information). Formalizing the metaphor of character constellations, these models incorporate the tools of network analysis to detect and analyze interrelations between characters. This is done by generating fictional social networks for each of the 170 novels based on co-occurrences of characters on the sentence level, an approach which is described in detail in chapter 3. Each chapter subsequently highlights an aspect of these fictional
social networks that affects the representation of social groups: their centrality (which types of characters are important or dominant), their communities (how characters are integrated or segregated into distinct groups), and their conflicts (the clashes between different social groups). Following recent developments within the field of cultural analytics (see section 1.3), these network analytical models are then used in the individual chapters to recursively go back and forth between statistical, pattern-based analysis and qualitative close readings of particular case studies from this corpus. As such, the results generated by the models developed throughout this book showcase what character-based critiques of literary representation gain by integrating data-driven methods into the practice of critical close reading.

By combining the formal methods of social network analysis with the interpretive tools of narratology in these models, this book thus opts for a data-driven critique of literary representation of which both measuring and reading are indispensable parts. The thesis it defends is that the literary representation of social groups is hierarchically structured along the axes of centrality, community, and conflict. This approach yields insights into the representation of social groups in a large body of texts from one sample year of Dutch literary production, and potentially as well into its changing dynamics throughout literary history, paving the way for future longitudinal research to extend the models and their findings presented here. Merging the critique of literary representation as evolving in the Netherlands from the 1990s onwards with more recent developments of data-driven approaches to culture within the field of cultural analytics, it hopes to contribute to a fruitful debate between research traditions and between quantitative and qualitative methodologies. In each of the chapters this is done by confronting the statistical, quantitative patterns with qualitative close readings of selected case studies from the corpus pinpointing which parts of the patterns are reflected or deconstructed.

1.1.1 Example: Character Constellations in Joost Zwagerman’s *De buitenvrouw* (1994) and Robert Vuijsje’s *Alleen maar nette mensen* (2008)

Discussions on two books of Dutch literary fiction published within a range of 14 years may serve to exemplify the relevance of this book’s theoretical and methodological approach. On a thematic level, Joost Zwagerman’s *De buitenvrouw* [*The mistress*] (1994) and Robert Vuijsje’s *Alleen maar nette mensen* [*Only decent people*] (2008) show striking resemblances. Both novels comment on
Dutch multicultural society by staging Dutch-born, male protagonists who have sexual desires for black female characters. Sensitive topics such as racism and colonialism are addressed through Theo Altena’s extramarital relation with the Dutch-Surinamese Iris Pompier in De buitenvrouw and through David’s search for the ‘intellectual negress’ in Alleen maar nette mensen. Although both novels could count on critical acclaim and/or institutional recognition, they received negative criticism for the ways in which black women are depicted. Most notably, public intellectual Anil Ramdas accused Zwagerman of a flat, stereotypical portrayal of his main character’s black mistress (1997). ‘We know what she smells like (nut oil and dewy woodland);’ he writes, ‘but not what she thinks.’ In response to Vuijsje’s novel, gender and ethnicity scholar Gloria Wekker contended that it seems hard for Dutch authors ‘to write about black women and to not associate them with sexuality’ (Meershoek, 2009). In turn, the ideological objections against Zwagerman’s and Vuijsje’s literary representation of black women were met with equally fierce arguments emphasizing the autonomous and therefore inviolable position of the literary author (Zwagerman, 1997; Van Aalten, 2009).

Although these discussions show a strong disagreement about the representation of the social group of black women as emerging from these two novels (critics such as Ramdas and Wekker emphasize their offensiveness, while the authors stress that these representations function as part of an autonomous literary-artistic endeavor), all those involved base their interpretations on the same words, sentences, and literary-stylistic configurations present in the texts. While both critics and authors stare at the same sky (novel), they discover different images in the patterns of the stars (characters) and interpret those images in light of their own norms, values, and beliefs. It is, furthermore, fully understandable that the authors and their critics ascribe different meanings to the patterns they see in these novels. As creators of these patterns, authors Zwagerman and Vuijsje might have stronger inclinations to defend the images of black women emerging from their novels as a crucial part of their literary-artistic vision. Conversely, the criticism on the depictions of black women by Ramdas and Wekker can partly be understood in light of their personal background (both were born in Suriname) and their political beliefs (both are known for their outspoken ideas on racial inequality in Dutch society). Without suggesting that one image is more correct than the other, the methodology developed in this book serves to guide such discussions on the representation of social groups by mapping out the concrete coordinates of characters, in order to provide an empirically informed blueprint of the character constellations present in the texts.
The following is what this approach could yield for the narrative worlds presented in Zwagerman’s *De buitenvrouw* and Vuijsje’s *Alleen maar nette mensen*. A cursory examination of these novels in terms of social networks of characters – or character constellations – helps to pinpoint some basic aspects of their depiction of social groups, of which this book highlights three: centrality, community, and conflict. In terms of centrality or importance, both Dutch-born, male protagonists – Theo (*De buitenvrouw*) and David (*Alleen maar nette mensen*) – are mentioned considerably more often than the black women they desire, and the frequency with which their thoughts and feelings are described stands in sharp contrast with what the readers get to know about the inner lives of the black female characters. The primary focus lies on the perspectives of Theo and David, which makes them the center of attention and makes it their stories and less those of Iris and Rowanda. In terms of the groups or communities in which the characters function, the novels are segregated by ethnicity and descent. In *De buitenvrouw*, Iris and Theo are part of the same community of teaching staff at a Dutch high school, but she remains an outsider as she is also part of a community of Surinamese migrants with language and norms exotic to the Dutch-born characters. An even stronger contrast is apparent in the communities in which David and Rowanda in *Alleen maar nette mensen* function: David was born and raised in an elite Amsterdam neighborhood where mostly rich people without a migration background live, whereas Rowanda lives in a black neighborhood full of criminal activity. The protagonists of both books make attempts to become part of these other, exotic communities, but this does not result in an integration between social groups. The segregation between these communities, finally, signals the conflicts and disparities between the Dutch-born, male protagonists on the one hand and the black female characters on the other. Both Iris and Rowanda witness racist remarks of people from the other community, and at the end of the novels they both exclude their Dutch love interests from their own community, resulting in antagonism and polarization between both represented social groups.

Such a blueprint of character constellations can serve as a point of departure for a discussion on the representation of social groups in these novels. In whichever way one might interpret the status of Iris in *De buitenvrouw* and Rowanda in *Alleen maar nette mensen* within the fictional societies depicted in these books, crucial to their centrality, their communities, and their conflicts in the narratives is how often and where they occur, and with whom they interact in which specific way. Although other elements also play a part, aspects of centrality, community, and conflict are co-constitutive of the hierarchical orders underlying
the novels’ representation of black women as opposed to white men. As literary scholar Philippe Hamon has noted, reading inevitably involves a process of hierarchization in which narrative elements such as characters are placed by the reader in a ‘value system’, privileging some and downgrading others (1984, p. 54). The claim of this book is that such hierarchies can be studied by both reading and measuring the textual elements these hierarchies are composed of. The literary representation of social groups contains an inevitable quantitative component; counting simple occurrences of characters can already provide a first indication of their status in the text. However, counting is meaningless without reading; it requires the interpretive act associated with it. The status of Iris and Rowanda is not reducible to their occurrences in the narrative; such counts have to be put in the broader thematic and stylistic context of the novel and the world it was produced in. For that reason, the present study goes back and forth between computer-assisted, quantitative analysis on the corpus of 170 novels as a whole and fine-grained, qualitative readings of particular case studies.

1.1.2 Delineation of the Study and Research Question

The contribution of the present study to the field of character-based critiques of literary representation is to develop an empirical, data-driven account of character constellations in the novelistic genre in combination with a narratological evaluation of these accounts. This book inquires into characters populating the Dutch literary field for the simple reason that I was trained in the literature of this language, but it should be possible, with a few language-specific adjustments, to apply its methods to the literature from any other language field. The focus lies on Dutch literature from the present day and age, and more specifically on a corpus of 170 Dutch books of literary fiction from 2012. Although literature from other time periods could also have been the object of focus, the current (Dutch) sociocultural climate provides an argument for studying today’s rather than yesterday’s literature. The increasing awareness of social, cultural, and economic hierarchies in a variety of societal domains invokes the question of how literature deals with these issues. It simply makes sense to study the representation of social groups in works of literary fiction produced in the same period that fueled attention to the hierarchical societal structures these groups are embedded in.

Case studies from this corpus are selected on the basis of two criteria: 1) Degree of conformation to or deviation from the statistics-based patterns generated by the data-driven models applied to the whole corpus of 170 novels. Novels that seem to fit these models perfectly showcase how the observed
Introduction

statistical patterns are illustrated on the micro-level of individual works, whereas novels that are outliers to the models demonstrate how individual works have the ability to nuance or deconstruct these statistical patterns. 2) Themes related to the topic of the chapter (centrality, community, conflict). Based on cursory examinations of novels that either conform to or deviate from the statistical patterns, works were selected for their depiction of characters from a certain gender, descent, education, or age as more or less central, as more or less belonging to a community, and as more or less engaged in conflict.

In the chapter on centrality, Özcan Akyol’s Eus (2012) is analyzed because its portrayal of characters with a migration background and female characters simultaneously conforms to and deviates from the model’s findings on how central these types of characters are in the corpus as a whole. In the chapter on community, both Philip Huff’s Niemand in de stad [Nobody in the city] (2012) and Mensje van Keulen’s Liefde heeft geen hersens [Love has no brains] (2012) demonstrate how the model’s findings on segregation and integration between groups is partly reflected and partly deconstructed through particular narrative mechanisms. In the chapter on conflict, the depiction of social class in Bart Koubaa’s De Brooklyncub [The Brooklyn club] (2012) provides a good example of how one-on-one conflicts between characters take shape within a single narrative and shows how the narrative strategies comment on the found statistical observation regarding such conflicts. Leon de Winter’s VSV, of daden van onbaatzuchtigheid [VSV, or acts of altruism] (2012) is used in the same chapter to explore how triangular conflicts (between three characters) contribute to a moral privileging of one social group over another, while Tommy Wieringa’s Dit zijn de namen [These are the names] (2012) deconstructs such a schematic opposition between groups through its narrative conflicts.

The question ‘How are social groups represented in present-day Dutch literature?’ forms the basis of this book. Underlying this question is the assumption that characters in novels constitute a fictional population that can be studied with the same tools as those used by social science scholars studying real-world populations. This assumption invokes age-old discussions dating back to Plato (The Republic, book 10, c. 375 BC) and Aristotle (Poetics, 335 BC) on the question of whether fiction is a mimesis, a reflection, of society – and more specifically, whether social structures are mirrored in literary fiction.

Eric Auerbach’s Mimesis: The Representation of Reality in Western Literature (1946/2003) is commonly regarded as the seminal twentieth-century study on the alleged mimetic power of literature. Roughly, this book follows Auerbach’s approach in the sense that it also foregrounds the relation between the artistic
dimensions of the literary work and the sociopolitical context it emerged from. In the current study, this ‘real-world’ context is rather specific: it targets the ways in which concrete social structures are represented in the social networks of characters in the literature of a specific period. There is, however, a radical methodological difference between this book and Auerbach’s approach. Whereas Auerbach generalizes about the literary work’s represented reality based on single, allegedly exemplary cases, the present study grounds its generalizations in data-driven observations within a larger, representative body of texts and evaluates the resulting statistical patterns through close readings of case studies. By analyzing the characters in the corpus as a fictional population with social networks similar to those of actual populations, this book furthermore explores and assesses the applicability of the sociological methods of network analysis to fictional narratives. In doing so, the book’s focus on the person-like qualities of characters, alongside their textual dimensions, also works to break a taboo in literary criticism: treating characters as if they were real people. Recently, Toril Moi traced back the historical roots of this taboo to L. C. Knights’ essay How Many Children Had Lady Macbeth (1933) and demonstrated that it is ‘intertwined with the promotion of a specific understanding of modernist aesthetics and with the belief that formalist analysis is the raison d’être of professional literary criticism’ (Moi, 2019, p. 61). Following Moi’s argument that this taboo evolved into an untenable dogma, the present study examines both the textual and the person-like dimensions of characters in order to enable a full picture of how groups of people are represented in narrative fiction. This study does not, however, make any claims as to answering the million-dollar-question of how narrative fiction reflects societal tendencies. In order to arrive at the contours of an answer to that question within the framework of the present research, a structural comparison between real-world social structures and fictional structures should be made. While applying the analytical tools used for studying real-world populations, this book focuses on the narrative fiction itself. The central research question of how social groups are represented in the fictional character populations in the corpus is examined through three general themes related to the structure of the fictional social networks, each of which forms a sub-question:

**Central question:** How are social groups represented in present-day Dutch literature?

**Sub-question 1:** How does the centrality of characters co-shape the representation of the social group(s) characters function in?
Sub-question 2: How do the communities in which characters function co-shape the representation of the social group(s) characters function in?

Sub-question 3: How do conflicts between characters co-shape the representation of the social group(s) characters function in?

Underlying each of the three sub-questions is the hypothesis that the fictional populations are structured according to a hierarchical order (cf. Woloch, 2003). The first question tests this most straightforwardly by exploring the various ways in which characters can be central, important, influential, or dominant in the narrative worlds they inhabit. The second and third questions build more indirectly on this hypothesis. By studying the communities in which characters function, insight can be gained into the degree to which the different social groups are either integrated with other character types in the fictional population or segregated into different clusters – who does and who does not belong to a specific group. Finally, the conflicts between characters illustrate the ways in which clashes between social groups produce hierarchies within the fictional population. The centrality, communities, and conflicts of characters are thus taken to co-constitute an answer to the question of how social groups are depicted in present-day Dutch literary fiction.

‘Social group’ is a rather broad term used in the social sciences to denote collections of people with similar social, cultural, or economic features. For the sake of clarity, this book focuses on only four of those features: gender, descent, education, and age. Not only are these features the most common variables in studies on actual social networks, they are also defining features of characters, as was first described by Aristotle in his Rhetoric (cited in Florack, 2010, p. 479). Subsequently, characters are categorized in the following analyses as functioning in social groups consisting of either male or female characters, characters either with or without a migration background, either higher or lower educated characters, and either older or younger characters.

Literary criticism focusing on such identity categories – the so-called ‘critique of representation’ – forms one of the two pillars of the theoretical framework underlying the analyses carried out in this book. The other pillar is constituted by the field of cultural analytics, a subfield of digital humanities in which data-driven methods are used to model culture and narrative quantitatively, often from a social or ideological perspective. The remainder of this chapter sketches the contours of the twofold theoretical framework in which the research is situated. Section 1.2 describes how the concepts of representation and ideology have been used in scholarly critiques of literary representation and makes clear
how these concepts are defined and operationalized in this book, as well as what the book's place is in the scholarly debate on Dutch literature in particular. In section 1.3, the data-driven aspect of the research is outlined by delineating how it fits within the debates on close and distant reading, particularly with regard to the notion of modeling as operationalized in the field of cultural analytics. The methodological background of the study is described in section 1.4, namely, the tools of narratology and network analysis, the corpus and dataset, and previous studies on the corpus and data. But before the theoretical and methodological frameworks are delineated, some basic questions have to be answered regarding the central analytical unit of this book: characters. What are we talking about when we talk about characters?

1.1.3 Character Studies

Fictional characters in novels are the primary point of focus of this book. Although the concept of character might seem ubiquitous and common sense (don't we simply know what characters are?), the vast number of studies disentangling its complexities suggests otherwise. As the topic has been studied in various disciplines and from a wide variety of theoretical angles, character studies have long been a dispersed area of research (Heidbrink, 2010, p. 67; Jannidis, 2013). Since the 1990s stronger trans- and interdisciplinary accounts of characters started to emerge with the publication of a range of monographs (Culpeper, 2001; Eder, 2008; Florack, 2010; Jannidis, 2004; Koch, 1991; Palmer, 2004; Schneider, 2000), indebted to the earlier theoretical work by scholars such as Uri Margolin (1983) and Shlomith Rimmon-Kenan (1983). From the 1990s onward, these scholars have shaped discussions on character. According to Henriette Heidbrink, these discussions consequently progressed via three lines of inquiry (2010, p. 72):  

1. How does the human-likeness of characters relate to the textual, visual, or auditory material from which they are made up of? Are characters in literary fiction predominantly linguistic or psychological entities?  
2. What is the meaning of characters in terms of their actions and functions?  
3. How does the materiality of characters (words, images, sounds) affect their reception by readers?  

Whereas the reception of characters (the third question) is not touched upon in this book, both the first and the second questions are addressed. The book’s
premise is that real-world social groups are represented through characters who are – to a greater or lesser extent – fictionalized linguistic representations of these groups.” Men, for instance, are first and foremost represented in novels through male characters. Psychological traits and cultural roles ascribed to men in a particular society might be reflected, criticized, or satirized in the male characters populating the novels written in that society. But there is more to characters than just their psychological, cultural, or social similarities to real people. As characters in novels are linguistic configurations, they are tied to the boundaries of the language from which they are made up of. Because of that, the literary representation of, for instance, men is not just a matter of studying male characters in isolation. The novel’s materiality in which these characters are embedded (the second question) is of equal importance: the thematic structure, stylistic devices, metaphors, and symbolism can all co-constitute the ways in which men are depicted. Moreover, gender representation is also shaped by the various ways in which both male and female characters act and function in relation to one another. The representation of Iris Pompier in Zwagerman’s *De buitenvrouw* (2008), for instance, is largely determined by her relation to protagonist Theo. Because the actions and functions of fellow characters in the narrative thus co-constitute the representation of a particular type of character, this book takes up a fundamentally relational approach to the study of characters by focusing on the social networks – the character constellations – they function in.

What is a character? Depending on the research question and area of interest, scholars have defined characters in a variety of ways. The ontological status of characters has been subject to philosophical debates: do they, for instance, only exist within or also outside the narrative in which they function (Reicher, 2010)? Or do characters primarily exist as imaginary beings in the minds of the audience (Culpeper, 2000)? Are characters ‘pieces of writing’ or ‘person like entities’, or a hybrid of those (Frow, 2014, p. 25)? While being fully aware of the ontological complexities surrounding characters, delving deeply into such questions is outside the scope of this book. Acknowledging both their textual and person-like dimensions, I will use the most common definition of characters as ‘fictive persons or fictional analogues to human beings’ (Eder et al., 2010, p. 7). As such, the book follows the definition of *The Living Handbook of Narratology*: ‘a text- or media-based figure in a storyworld, usually human or human-like’ (Jannidis, 2013). A consequence of adopting this definition is that nonanthropomorphic beings, such as animals or inanimate objects, are not taken into account in this study, although present-day novels sometimes – but rarely – feature nonhuman-like characters. Another reason for using this definition
stems from the book’s focus on the representation of social groups. Although moral rights can be ascribed to wolves, dragons, and flying pancakes, they are usually not considered as social groups in the research disciplines of the social sciences.\textsuperscript{19} As this book aligns itself methodologically with the social sciences through its use of social network analysis, it defines social actors in literary texts in strictly anthropomorphic terms. It is, furthermore, hard if not impossible to ascribe demographic labels to nonhuman-like characters (e.g., is the big bad wolf in ‘Red Riding Hood’ higher or lower educated?).

Characters consist of different properties and dimensions. Prototypical properties distinguishing characters from other types of objects or entities in the narrative relate to their mental interiority; characters possess ‘mental states, such as perceptions, thoughts, feelings, and aims’ and have ‘both an outer appearance and an inner state of the psyche that is not visible from the outside’ (Eder et al., 2010, p. 13). Apart from their physical and psychological properties, another property of characters is that they typically perform certain social roles in the narrative (ibid.). All these properties of characters consequently function within different dimensions. In \textit{Reading People, Reading Plots} (1989), James Phelan makes a useful and frequently cited distinction between three dimensions:

1. The synthetic dimension relates to the artificiality of characters and comprises all the narrative elements out of which it is constructed (p. 2).
2. The mimetic dimension denotes the character’s relation to recognizable human traits (ibid.).
3. The thematic dimension relates to what a character stands for, what it represents (e.g., a social group) (p. 3).

In each of this book’s chapters, the synthetic dimension of characters – the words which they are made up of – is analyzed both quantitatively and qualitatively in order to gain insight into their thematic dimension, that is, how they are representative of one or more social groups. The mimetic dimension is presupposed in the analyses: features of characters relating to the social groups in which they are embedded are implied by both their synthetic and thematic dimensions.

1.1.4 Characterization and Character Types

How to determine to which social groups a character belongs, and how to recognize traits related to these groups? This question relates to the mechanism of characterization and the different character types resulting from specific forms
of characterization. Although often the gender, descent, education, and age of characters can be deduced relatively easily from their narrative context, the values associated with these demographic features arise from the specific ways in which characters are characterized. In its broadest sense, characterization can be defined as ‘the process of connecting information with a figure in a text so as to provide a character in the fictional world with a certain property, or properties, concerning body, mind, behaviour, or relations to the (social) environment’ (Eder et al., 2010, p. 32). Such properties can be ascribed to characters through direct characterization: e.g., ‘She – blond hair and blue eyes [body] – was smart and arrogant [mind], and tended to be rude to others [behavior + relations]’. Or through indirect characterization: e.g., ‘Her reading of Nietzsche’s Also Sprach Zarathustra in the local pub might evoke a peculiar kind of awe in the people obsessively staring at her appearance’ [mind + behavior + relations + body]. Such processes of characterization shape the social, economic, and cultural features of a character.

The best-known distinction between modes of characterization comes from novelist and critic E. M. Forster, who distinguishes in Aspects of the Novel (1927) between round and flat characters. In his view, round characters show more narrative development and are made up of multiple, sometimes changing features, whereas flat characters ‘are constructed round a single idea or quality’ and ‘when there is more than one factor in them, we get the beginning of the curve towards the round’ (Forster, 1927, p. 103). How to decide if a character is flat or round?

The test of a round character is whether it is capable of surprising in a convincing way. If it never surprises, it is flat. If it does not convince, it is a flat pretending to be round. (ibid., p. 118)

As the criterion of ‘surprising in a convincing way’ is highly vague, it very much remains a matter of interpretation whether or not to categorize a character as either round or flat. There are, however, immediately recognizable instances of both types. An example of a round character in contemporary popular culture is Frodo in The Lord of the Rings trilogy. From the quiet peace of the shire to the fires of Mount Doom, the multifaceted hobbit goes through several stages of personal development. A huge gap lies between the first encounter with Frodo in the shire and his destroying of the ring at the end of the trilogy. Frodo undoubtedly witnessed genuine spiritual growth, and his spiritual journey could arguably be labeled as surprising.
An example of a flat character from the same trilogy is the figure of the orc: the 'single idea or quality' (ibid., p. 103) represented by orcs is evil or wickedness. The personal history of orcs is not genuinely articulated, nor do they develop into something differently – they are first and foremost interchangeable evil and wicked creatures. Although flat characters tend to be negatively connotated in contemporary criticism, Forster saw two advantages for a novel writer to stage them: because they stay the same throughout the narrative, they are both easily recognized and remembered by the reader (ibid., p. 105). These negative connotations, however, are due to the fact that flat characters tend also to be caricatures or stereotypes.21

Stereotypes are often used as arguments in ideological discussions on representation, which is the case in the discussions on Zwagerman’s De buitenvrouw and Vuijsje’s Alleen maar nette mensen (see section 1.1.1 of this chapter). In the qualitative parts of the analyses presented in the subsequent chapters, stereotypes are, furthermore, used as possible indicators of the status of a character in a narrative. What, then, is a stereotype? The present-day usage of the term ‘stereotype’ as a simplified, reductionist form of characterization of a collection of people or of a set of practices dates back to Walter Lippmann’s description of the term in his book Public Opinion (1922). Whereas today stereotypes are mostly associated with rather negative, and even offending, figures of speech, he primarily emphasized the usefulness of stereotypes in everyday life. Lippmann describes processes of stereotyping as a natural effect of the way in which human beings perceive the world:

For the most part we do not first see, and then define, we define first and then see. In the great blooming, buzzing confusion of the outer world we pick out what our culture has already defined for us, and we tend to perceive that which we have picked out in the form stereotyped for us by our culture. (Lippmann, 1922)

Although this definition lacks the strong moral dimension the term has today, Lippmann also sees stereotyping as a simplified, reductionist process, as people, in his view, fall back on ‘what our culture has already defined for us’. When people obey a natural tendency to ‘define first and then see’, they do not have to decide for themselves. Etymologically, the term dates back to premodern printing practices, in which stereotypes referred to printing plates that could be used over and over again.22 As such, stereotypes made the printing process more time-efficient and cost-effective. Similar to Lippmann’s use of the term, this particular historical context foregrounds aspects of non-originality and
repetitiveness. In order to make their work easier, printers made use of already available, reusable printing plates instead of using new, original printing plates each time. In Lippmann’s view, stereotypes make life easier because people can rely on repetitive images instead of creating new images every time. In a similar vein, Forster values the pragmatic aspects of a flat character as being easily recognizable and remembered (1927, p. 105). There is, in short, a certain convenience, and perhaps even necessity, to stereotypes.

More recent scholarly work on cultural stereotypes also acknowledges their simplified, reductionist, repetitive nature, but it tends to put more stress on the ideological problems related to stereotypical representations (e.g., Florack, 2010; Gymnich, 2010; Schweinitz, 2010).23 In the seminal article ‘The Role of Stereotypes’ (1999), Richard Dyer distinguishes between the social type and the stereotype. Both character types are easily recognizable in narratives and thus tend to be rather flat. There is, however, an important difference: social types refer to ‘those who “belong” to society’, whereas stereotypes refer to ‘those who do not belong, who are outside of one’s society’ (Dyer 1999). Who does and who does not belong is indeed a matter of perspective. In present-day Dutch novels, a character pertaining to, for instance, the social type of the Dutch character can be introduced by just a few signals (born in the Netherlands, conforming to certain Dutch customs, norms, and values). But such a social type can be used in the plot ‘in a much more open and flexible way than can stereotypes’ (ibid.). They can fulfill almost every role in the narrative, which is illustrated by the fact that social types of Dutch characters are prevalent in a wide variety of characters ranging from Gerard Reve’s Frits van Egters to Arnon Grunberg’s Tirza.

Conversely, the stereotype of, for instance, the Muslim extremist character is bound to a specific set of functions: they are radical, evil, dangerous. In the fourth chapter of this book, on community, the stereotypical representations of Muslim extremist characters in Leon de Winter’s VSV (2012) illustrate how such stereotypes ‘maintain sharp boundary definitions, to define clearly where the pale ends and thus who is clearly within and who clearly beyond it’ (ibid.). In this definition, stereotyped characters are ideologically restricted to a very specific, one-dimensional depiction that does not do justice to the full complexity of the social group they are representative of. Determining whether a character is a flat or a round character, a social type or a stereotype is thus a convenient way to gain insight into the one- or multidimensionality ascribed to the social group(s) a particular character represents. In the qualitative parts of the analyses in the subsequent chapters, this is done through narratological analyses of the ways in which specific characters are narrated, focalized, and characterized.
1.1.5 The One Versus the Many

In the last decade or so, quantitative and data-driven accounts of characters started to emerge (Bamman, Underwood, & Smith, 2014; Jockers & Kirilloff, 2016; Kraicer & Piper, 2019; Piper, 2018; Underwood, Bamman, & Lee, 2018). Either indirectly, or directly in the case of Piper (2018, p. 219), these studies respond to a central claim of Alex Woloch’s book *The One vs. the Many: Minor Characters and the Space of the Protagonist in the Novel* (2003). ‘Narrative meaning’, Woloch contends, ‘takes shape in the dynamic flux of attention and neglect toward the various characters who are locked within the same story but have radically different positions within the narrative’ (p. 2). Woloch makes something explicit that might seem quite obvious but has lacked scholarly attention: there is a huge discrepancy in the distribution of attention for characters in a narrative. Most attention is devoted to the protagonist, with only sparse attention left for a wide range of minor characters. These discrepancies raise fundamental questions regarding the hierarchical structure of narrative representation:

> How can many people be contained within a single narrative? How do different narrative forms accommodate the surge of many people into a single story? How do they encapsulate and convey the impact of a human being – of varied human beings – within a coherent literary structure? In these questions we can see the outline of a different, almost inverted, history: a history that would trace not how the literary form, in its intricate coherence, is rendered into a living organism, but how living persons get rendered into literary form. (p. 11)

By formalizing and quantifying the concept of character, the data-driven character studies listed above provide an empirical account of the ways in which ‘many people [can] be contained within a single narrative’. Counting characters and their interrelations, in combination with close readings of those character constellations, arguably results in a more complete view on the dominance, subordinance, and hierarchies of characters than a solely qualitative assessment of texts. In line with these quantitative approaches to the study of characters, this book builds forth on the recent attempts to model the relation of ‘the many’ to the ‘one’. By means of computational, data-driven analysis, it broadens the scope from the main character(s) to all identifiable characters in the corpus. Combining quantitative and qualitative analysis in each individual chapter, it aims to gain insight into the ways in which living persons are rendered into literary form while paying special attention in the qualitative parts of the analyses to the
different (round/flat, social/stereotypical) character types they are represented through.

1.2. CRITIQUE OF REPRESENTATION

1.2.1 Representation and Ideology
In this book, critique of representation refers to the study of cultural and literary representation as popularized by scholars such as Edward Said, Stuart Hall, and, in the Netherlands, Mieke Bal and Maaike Meijer. The intellectual tradition of this field of inquiry forms the main theoretical background of the analyses carried out in the subsequent chapters. More specifically, the present research builds forth on the ways in which this tradition has conjoined representation and ideology as concepts in the analysis and interpretation of narrative fiction. In the seminal book *Representation* (1997), Stuart Hall systematically explores the notion that ‘languages work through representation’, where language is not narrowly understood as Dutch or English but more broadly as a ‘dialogue’ between people through which meaning is communicated (p. 4). In a similar vein, Maaike Meijer applies insights from feminist and postcolonial criticism to the study of cultural representation (1996). Central to both of these books is an operationalization of poststructuralist theory and an application of ideas from the ideological strands of Marxist thought, gender studies, and postcolonial studies to the analysis of cultural representation.

What is representation? In the classic accounts of literary theory by Plato and Aristotle, artistic expressions such as literature are viewed as a representation of reality. For Plato, this was a reason to ban literature from his ideal of the perfect state: as a potentially illusory reflection of reality, literature might lead people astray from the path of truth. In modern times, the idea that literature has a reflective – a mimetic – component is still vivid, although scholars have problematized the ways in which the mediation between literature and reality takes shape. Building on the mediating aspect of representation, Maaike Meijer distinguishes ‘representation’ from the semiotic concept of ‘sign’, as the first does and the latter does not require a human mediator. Arguing that ‘representation is always of something or someone, by something or someone, to someone’ (1995, p. 12), W. J. T. Mitchell characterizes representation as a triangular
configuration between the represented, the representer, and the audience for whom the representer creates the represented. In its simplest form, the three components of representation in the context of this book are the actual social groups of people as represented objects, the literary representations of these social groups as the representers, and the readers as the targeted audience.

For a critique of representation, each of these components requires a different approach. For instance: the represented social groups can be studied through anthropological analysis, the literary representations of these groups can be studied through text analysis, and a reception analysis of the texts can be carried out with the readers as subjects of study. Without denying the importance of both the actual social groups depicted in the representations and the readers, this book studies the text representations (the representers) largely in isolation from these other two components of representation. Although the literary representation of social groups in the corpus is co-shaped by the features and actions of the respective social groups and its reception by readers, the present text-centric approach aims to foreground the structure and hierarchies within the fictional populations as a starting point for any future analysis considering these other two dimensions of representation. By doing so, the research aligns itself in the first place with formalist approaches to representation deemphasizing both the represented objects (the actual social groups) and the audience (the readers), and stressing the materiality of the literary objects (the literary depictions of the social groups). But while prioritizing the literary object as representer, it does not go as far as to assume that ‘literature is about itself’ and that ‘novels are made out of other novels’ (Mitchell, 1995, p. 16). The book’s prioritization of the text is not the result of a conviction that the representation of social groups functions in isolation from the actual represented social groups and the readers perceiving these representations, but is due to a methodological choice to narrow down the focus to one of these three dimensions. By means of focusing on this dimension, the study aims to show that blueprints of character constellations at the level of the text – how representations of social groups are rooted in the words on the pages – provide a solid starting point for any discussion on literary representation.

Quite obviously, processes of representation are the objects of critiques of representation. Less obvious, perhaps, is that these critiques tend to be deeply involved with ideological perspectives on these processes. For Mitchell, it is only logical that ideology is a core component of this analytical tradition:
It should be clear that representation, even purely ‘aesthetic’ representation of fictional persons and events, can never be completely divorced from political and ideological questions; one might argue, in fact, that representation is precisely the point where the questions are most likely to enter the literary work. If literature is a ‘representation of life,’ then representation is exactly the place where ‘life,’ in all its social and subjective complexity, gets into the literary work. (Ibid., p. 15)

With regard to the literary representation of social groups, it is immediately apparent that such ‘political and ideological questions’ are at play. An author can choose to foreground specific features of characters more than other features: should they stress the physicality, the psychology, the material possessions, the morality, or other dimensions of the respective characters? For instance: an author who – (sub)consciously – chooses to characterize female characters more emphatically by their bodily traits, and to conversely stress the intellectual capabilities of the male characters, is not making a neutral choice. Ideologically, this author creates associations between femininity and physical appearance, and between masculinity and psychological interiority. Based on these associations, further value-laden connections can arise in the narrative, such as those between women and eroticism or sensuality, and between men and rationality. Such textual connections have an ideological component because they co-shape conceptions of the male and the female.

For critiques of representation it thus seems inevitable to inquire into the ideological dimensions of cultural representations. How, then, should ideology be understood as a concept within the framework of this book? Like the concept of representation, ideology is one of the key terms in literary and cultural studies. Since its coinage by the French philosopher Destutt de Tracy in Éléments d’idéologie (1827/2015) to propose a new science of ideas, the term has witnessed a complex history. A closer look at the history of the term shows that it has denoted a wide variety of different, often conflicting meanings (Eagleton, 1991, pp. 1–31). This book follows a text-centric definition of ideology as ‘the frame of values informing the narrative’ as put forward by narratologists Luc Herman and Bart Vervaecck in The Living Handbook of Narratology (2013). More specifically, ideology in narrative fiction is considered to ‘[install] hierarchical relationships between pairs of oppositional terms such as real vs. false, good vs. bad, and beautiful vs. ugly’ in the text (ibid.). The above hypothetical example on male-female associations demonstrates that the ideological aspects of representation can foster oppositions such as women as bodies versus men as minds.
Marxist theory has undoubtedly left a major mark on the use of ideology in contemporary literary and cultural criticism. Initially, Marx and Engels used the term to refer to ‘a widespread form of epistemological error that a new, more empirically-based – sometimes called “scientific” – mode of thinking could avoid, if not eventually abolish’ (Kavanagh, 1995, p. 310). In this definition, ideology is the opposite of truth and pertains to false beliefs about reality. More recent Marxist criticism by Louis Althusser does not equate ideology with false beliefs but rather sees it as the means through which the relation between culture and politics is shaped (1970; 1971). In this view, ideology is an inherent aspect of each cultural product because it functions as a ‘system of representation’ arising from the ‘social process that works on and through every social subject, that, like any other social process, everyone is “in”, whether or not they “know” or understand it’ (Kavanagh, 1995, p. 311).

Studies of ideology in narrative fiction are divided by Herman and Vervaeck into psychoanalytic, sociological, and discursive approaches (2013). Building on the work of Freud and Lacan, psychoanalytic approaches tend to focus on the way in which the reader creates ideological connections in their mind (e.g., Davis, 1987). From a Marxist perspective, sociological approaches stress the ideological aspects of the social, historical, cultural, and economic contexts in which narratives function (e.g., Williams, 1977). Discursive approaches center on the ideological constellations contained in the literary texts themselves (e.g., Bakhtin 1935/2003; Hamon, 1984). While the psychoanalytic approach is outside the scope of this book, both the sociological and the discursive approach have a place in it. In terms of representation, this study focuses primarily on the literary representations of social groups in the texts themselves and thus aligns itself in the first place with the discursive approach to ideology in narrative fiction. But while it does not study either readers (psychoanalytic approach) or the actual represented social group (sociological approach), it does hinge on the idea that the characters populating the novels can be studied as if they were real people, following Toril Moi’s recent proposition to go beyond this taboo in literary criticism (2019, pp. 27–75, see also section 1.2 of this chapter). A study on how fictional characters represent actual groups of people cannot rely on an analysis of the textual components of characters only, but requires one to make a connection between real people and the person-like qualities of characters that are representative of these people. Using the tools of social network analysis, the book thus contextualizes the fictional characters in the texts as representatives of actual social groups. In this specific sense, it hopes to invoke a dialogue between
Introduction

the discursive, text-centered approach and the sociological, contextual approach to ideology in narrative fiction.

While ideology is discussed above in a general sense, critiques of literary representation often focus on specific ideologies, values, or belief systems, building on one or more traditions of criticism, such as gender, postcolonial, queer, disability, or mad studies. Gender and ethnicity have demonstrated to be two of the most widely used identity categories targeted in critiques of representation. Influenced by feminist theorists such as Judith Butler (1990) and postcolonial theorists such as Edward Said (1978), hierarchies between genders, sexes, or sexualities, and between ethnicities, races, or descents have been a central topic in the study of literary and cultural representation. The prevalence of feminist and postcolonial approaches to representation in literary texts is exemplified by the emergence of separate branches of narratology centering on gender and ethnicity. As one of the main proponents of feminist narratology, Susan Lanser contends that the branch’s existence is based on ‘the shared belief that sex, gender, and sexuality are significant not only to textual interpretation and reader reception but to textual poetics itself and thus to the shapes, structures, representational practices, and communicative contexts of narrative texts’ (2013). Similarly, the emergence of postcolonial narratology is motivated by the belief that ethnicity, race, and descent are crucial to how narratives are organized. In Gerald Prince’s view, a postcolonial narratology should be ‘sensitive to matters commonly, if not uncontroversially, associated with the postcolonial (e.g., hybridity, migrancy, otherness, fragmentation, diversity, power relations)’ and should ‘[envisage] their possible narratological correspondents and [...] [incorporate] them’ (2005, p. 373).

Such efforts to develop gender and postcolonial narratologies, among others, exemplify the need for analytical models to specify where and how values on, for instance, gender and race are expressed in texts. The work of scholars such as Susan Suleiman (1983), Philippe Hamon (1984), Liesbeth Korthals Altes (1992), and Vincent Jouve (2001) showcases how ideological dimensions can be traced back to the formal characteristics of texts. While these scholars attribute a greater or lesser importance to the role of the reader in the emergence of ideology in literature, their work provides inspiring examples of how the ‘ideology effect’ or ‘value effect’ is rooted in concrete words on the page. An important role is ascribed to characters: which norms and values are imposed by a text can partly be deduced from the actions, utterances, thoughts, feelings, and perceptions of characters. Characters, in this narratological tradition, are often seen as carriers of ideology. How something is narrated or focalized by a character is expressive of
the values represented by its point of view. In *Poétique des valeurs* (2001), Vincent Jouve describes the ways in which a text conveys values through characters, as well as through other narrative elements. In his view, values are located in specific points in the text – ‘points-valeurs’ – and are manifested in characters’ thoughts, words, and actions (pp. 35–88). The values characters express through thinking, talking, and acting are subsequently evaluated by the text explicitly (e.g., judgements on a character), through characterization (e.g., which social, cultural, or economic features are ascribed to a character), through modes of narration or focalization (e.g., we might be inclined to empathize more strongly with a first-person narrating character because we are close to their perspective), or through ‘mise en texte’ (e.g., in which geographic spaces the character functions). Such narrative mechanisms are taken by Jouve to co-shape the ‘value effect’ of the text or the ideology it imposes on its reader.

A concrete example of how ideology is manifested in the structural features of literary works is provided by Susan Suleiman’s *Authoritarian Fictions* (1983). In it, she develops a range of narratological models to study the novelistic genre of the ‘roman à thèse’, which she defines as ‘a novel written in the realistic mode (that is, based on an aesthetic of verisimilitude and representation), which signals itself to the reader as primarily didactic in intent, seeking to demonstrate the validity of a political, philosophical, or religious doctrine’ (p. 7). Novels in this genre convey ideology in the form of a particular message, statement, doctrine, belief, norm, or value. As such, the *roman à thèse* ‘seeks not only to impose a single meaning, but to propose a system of values’ (p. 56). Suleiman analyzes how such novels invoke an ultimately unambiguous and dualistic ‘system of values’ by dividing narrative elements such as characters, events, and places into a positive and a negative pole. A character-based example: a text can convey a certain message, or ideology, by a positive portrayal of character X voicing this particular message and a negative portrayal of character Y voicing the opposite of this message. Throughout her book, Suleiman describes the defining traits of this genre by breaking down the narratives by means of formal schemata and almost algebraic formulas. In doing so, she anchors the ideological dimension of these texts in the words on the page, an approach followed in the analyses carried out in the subsequent chapters of the present study.

On a more general level, the ideological dimensions of texts, and representations of gender, ethnicity, or any other identity category in particular, are often critiqued by examining how ‘the Other’ is represented in opposition to the dominant perspective adopted in the text. While it is unclear who coined the term, it can be traced back to Georg Wilhelm Hegel, who ascribes a constitutive
role to the Other in the formation of self-consciousness and self-identity in Phänomenologie des Geistes (1807/1832). In literary studies, the term is nowadays most notably associated with Edward Said’s seminal Orientalism (1978), through which it has come to be understood as the process through which entities such as characters are represented as an alterity to the central perspectives, the dominant characters, or the common worldviews as presented in the text. Such processes of othering are closely tied to mechanisms of characterization as described in section 1.1.4 of this chapter; the flatness or roundness of a character tends to influence its position in the narrative. Narratological models such as those of Suleiman (1983) and Jouve (2001) can, furthermore, help to pinpoint where otherness resides in the systems of values imposed by texts. In the qualitative parts of the analyses presented in the subsequent chapters, narratological analysis is used to trace where and how particular represented social groups are depicted as Other or Self, marginal or central, subordinate or dominant, silent or present.

1.2.2 Critiques of Representation in Dutch Literature

Published in 1991, De canon onder vuur [The canon under fire], edited by Ernst van Alphen and Maaike Meijer, symbolizes the institutionalization of ideologically oriented critiques of representation within the field of Dutch literary studies. The book’s subtitle, ‘Reading Dutch literature against the grain’, invokes a rather oppositional stance toward the object of inquiry. In order to expose ‘the less pleasant smells’ of canonical Dutch literary texts, each contribution addresses the – amongst others – sexist and racist dimensions of individual texts operating in ‘a realm of sacredness’.

In the decades following this book, a wide range of critiques of representation in its spirit have been published. The following three examples demonstrate how this line of research tends to focus on the narratological units of characters in order to lay bare processes of stereotyping and othering in Dutch literature. Analyzing representations of gender and race in De stille kracht [The hidden force] (1900) by Louis Couperus, Pamela Pattynama emphasizes that ‘the differences between the colonized and the colonizers [in the novel] are often rendered metaphorically as the unfathomable secrets and dangers that enveloped the Dutch colonial community of the East Indies’ (1998, p. 84). In line with the critical scholarship on Conrad’s Heart of Darkness (see section 1.2.1), Pattynama describes how Couperus depicts the Dutch colony as a mysterious place full of secrets and dangers, and how he portrays the colonized characters as alien forces embodying those secrets and dangers. Such strategies of othering are also
the point of focus in Liesbeth Minnaard's analysis of *Blank en geel* [White and yellow] (1894) by Lodewijk van Deyssel. Focusing on the encounter between a young Dutch bourgeois woman and a Chinese merchant, Minnaard shows 'how Van Deyssel’s novel tells the story of a dissident, improper desire for an exotic other, and reflects on exotic presence within the Dutch centre’ (2010, p. 75). In the eyes of the Dutch character, the Chinese merchant is an ‘Oriental prince’ and a ‘mysterious, almost supernatural being’ (p. 71). From a gender perspective, Maaike Meijer analyzes the representation of male and female characters in sex scenes by Jan Wolkers (Meijer, 1996). In her analysis, she observes traces of the Other in the ways the female antagonists in Wolkers’s work are presented as naive, sexually compliant, and submissive.

*De canon onder vuur* and the studies published in its wake mark the transition from a structuralist to a poststructuralist paradigm in Dutch literary studies. As iconic figures of Dutch structuralism, J. J. Oversteegen (1965) and A. L. Sötemann (1966) left their mark on postwar text-centric scholarship on Dutch texts. Especially, Sötemann’s dissertation *De structuur van Max Havelaar* [The structure of Max Havelaar] (1966) forms a classic example of a text-centric study dissecting the structural elements of one, highly canonical, novel in order to reconstruct its ‘meaning’. In line with Roland Barthes’s early structuralist theorizations (Barthes, 2006), Sötemann read Multatuli’s novel exclusively in terms of the elements present in the text without paying attention to any contextual information. Adopting the stance of an ‘ideal reader’, he considered it ‘necessary to devote careful attention to the extraordinary qualities of this work’ (Sötemann, 1966, p. 5). The premises of the scholarly work carried out in the tradition of Sötemann and those of the critiques of representation kick-started by *De canon onder vuur* could hardly be further apart. Whereas an outright admiration for the structure and meaning of canonical literary works is apparent in the first, the latter adopts a fundamentally resistant stance toward the text.

Resistant reading as practiced in these Dutch critiques of representation is closely related to a poststructuralist, deconstructivist strategy to break open the text in order to illustrate its ambiguity and indeterminacy (cf. Culler, 1983). In the structuralist framework of the likes of Sötemann, such deconstructivist reading strategies conflict with the ways in which the structuralist admiration for canonical texts motivates the reconstruction of structures and meanings instead of demonstrating the openness and multiplicity of texts. The opposition between these two dispositions toward the text is best understood by the terms ‘hermeneutics of admiration’ and ‘hermeneutics of suspicion’ as popularized by
Rita Felski in her essay ‘After Suspicion’ (2009). In it, she builds forth on Paul Ricoeur’s observation that recalcitrant thinkers such as Marx, Nietzsche, and Freud have shaped scholarly modes of suspicious interpretation (1970). Signaling the pitfalls and merits of both the hermeneutics of suspicion and the hermeneutics of admiration, Felski proposes a mode of ‘reflective reading’ moving beyond the suspicion-admiration-dichotomy, which later became known as ‘post-critique’.40

The present book takes a specific position within the scholarly tradition of critiques of representation on Dutch literature. It is aligned with the deconstructivist, resistant, suspicious readings in the spirit of *De canon onder vuur* as it aims to unravel representations of social groups in present-day Dutch literature by analyzing, among others, narrative mechanisms of othering and stereotyping. Its assumption that such narrative mechanisms are a central part of the literary representations under scrutiny is in line with the work of Maaike Meijer, Pamela Pattynama, and Liesbeth Minnaard as showcased above as examples of this tradition.

Although the book does not in any sense conform to the admiration of canonical works as apparent in the work of structuralists such as Sötemann, its primarily discursive approach to literary representation is indebted to their strong focus on the structure of literary texts. While, other than Sötemann, it does underwrite the belief that the meaning of literature is constituted by more than the text itself, its statistical analyses nevertheless isolate the texts from their institutional contexts and their readers.41 Focusing on a larger body of texts than the Dutch structuralists did, its aim is to lay bare structural patterns and trends of representation based on primarily textual elements. However, contrary to structuralist attempts to reconstruct a definitive meaning from texts, the narratological evaluations of these statistics-based textual patterns are then used in the qualitative parts of the chapters to point at the ambiguities and indeterminacies of representation. This is especially relevant in light of the rise of digital and empirical methods in literary studies. As Lucas van der Deijl has pointed out, data-driven studies in the humanities tend to rely (implicitly or explicitly) on structuralist notions without reflecting on the hermeneutics of algorithms (2015, p. 49). In a seminal article in which the term ‘computational turn’ is coined, David Berry highlights the need for such a hermeneutics of algorithms (or code), contending that ‘understanding the digital is in some sense also connected to understanding of code through study of the medial changes that it affords, a hermeneutics of code’ (2012, p. 6).

Triangulating an ideologically oriented critique of representation, a structural textual analysis, and a hermeneutics of algorithms, the book aims to transcend
the opposition between the deconstructivist, resistant accounts of literary representation and structuralist approaches to text analysis. By incorporating data-driven, empirical methods, it shows how text-centric, statistical analyses can help to gain insight into recurring patterns of literary representation. In turn, these structural textual patterns are then deconstructed qualitatively in each chapter by determining the moments at which individual texts conform to or deviate from them. As narratological evaluations of the statistical, text-centric patterns, these close readings thus seek to resist – to be suspicious of – the found structures and the logic of the algorithms developed for this research. The next section dives deeper into the merits and pitfalls of algorithmic approaches to literature.

1.3. CULTURAL ANALYTICS

1.3.1 Debates on Distant versus Close Reading

In the last decade or so, ‘digital humanities’ have increasingly become a buzzword in academia. Rumor has it that scholars applying for a grant maximize their chances of success by mentioning digital humanities as a component of their future research. The field is witnessing a rapid institutionalizing, as humanities faculties all around the globe have started to incorporate digital humanities courses, minors, specializations, and master programs into their curricula. Such developments might give the impression that digital humanities is the new, popular kid on the block. But in fact, data-driven humanities research has been around for at least 80 years.42

Although it seems common sense to juxtapose the ‘soft’ humanities disciplines with the ‘hard’ natural sciences, speculative interpretation with factual analysis, words with numbers, intricate historical interrelations exist between the notions of literacy and numeracy. In a contribution to Defining Digital Humanities (2013), Edward Vanhoutte offers a detailed account of the history of these interrelations.43 He sees the birth of the digital humanities symbolized in the person of Ada Lovelace (1815–1852), daughter of the most literate poet Lord Byron and the highly numerate mathematician Anabella Milbanke. In Notions sur la machine analytique de Charles Babbage (1842), Lovelace ruminates about using Charles Babbage’s (1791–1871) Analytical Engine for more than just
mathematical calculations,\textsuperscript{44} such as for automatically producing music. From another angle, Jesuit priest Roberto Busa (1913–2011) is commonly regarded as the father of modern data-driven humanities research. In the early 1940s, Busa started working on the \textit{Index Thomisticus}, a searchable database of all the works of Thomas Aquinas.\textsuperscript{45} From the 1950s onward, he started cooperating with IBM to automate certain aspects of this database (Vanhoutte, 2013, p. 127). This cooperation between old-school humanist Busa and the IBM technology company is usually regarded as a symbolic stepping-stone toward the practical integration of digital technology into the humanities disciplines.

Since the second half of the twentieth century, a wide variety of terms have circulated emphasizing different aspects of what nowadays goes under the term ‘digital humanities’. The term ‘humanities computing’ was used from the 1950s, and more regularly from the 1980s, to denote ‘computing in the humanities’ as opposed to ‘computing for the humanities’ (p. 140).\textsuperscript{46} With its first issue published in 1966, the academic journal \textit{Computers and the Humanities} functioned as platform for research carried out under the header of humanities computing. It was only since the publication of \textit{A Companion to Digital Humanities} in 2004 that ‘digital humanities’ was commonly accepted as an umbrella term for research combining digital technology with one of the humanities disciplines (Nyhan, Terras, & Vanhoutte, 2013, p. 2). The metaphor of a big tent is often used to explain the broad nature of the term: ‘Digital Humanities as a term does not refer to such a specialized activity, but provides a big tent for all digital scholarship in the humanities’ (Vanhoutte, 2013, p. 144). As the term is so generic that it obscures rather than elucidates the content of the research, the alternative term ‘cultural analytics’ will be used to characterize the data-driven component of the research carried out in this book (more on this in section 1.3.2).

In 2011 David Berry signaled the emergence of a computational turn in the humanities. Other than earlier uses of digital technology in humanities research, recent studies appeared questioning ‘the “hard core” of the humanities, the unspoken assumptions and ontological foundations which support the “normal” research that humanities scholars undertake on an everyday basis’ (p. 4). According to Berry, this new wave of digital humanities did not just use digital technology as an auxiliary to existing practices but ‘point[ed] the way in which digital technology highlights the anomalies generated in a humanities research project’ and led to ‘the questioning of the assumptions implicit in such research, e.g. close reading, canon formation, periodization, liberal humanism, etc.’ (ibid). The questioning of these assumptions subsequently led to heated debates on the status of quantity, measurement, objectivity, and replicability in the humanities.\textsuperscript{47}
In these debates, the digital humanities are often explicitly framed in opposition with the ‘traditional’ humanities.

In literary studies, the juxtaposition between the ‘new’, computer-oriented approach and the ‘traditional’ approach is most clearly exemplified by the work of Franco Moretti and Matthew Jockers. Coining the terms ‘distant reading’ (Moretti, 2013) and ‘macro analysis’ (Jockers, 2013) as alternatives for ‘close reading’ and ‘micro analysis’, both scholars quite unfortunately contributed to a polemical, and often unproductive, opposition between methodologies. In his account of the merits and pitfalls of a macro analysis of literature, Jockers asserts that micro analyses are commonly based on what he pejoratively calls ‘anecdotal evidence’ (2013, p. 5). In a similar vein, Moretti criticizes the inevitable subjectivity of close reading and notoriously contends that data are ‘independent of interpretation’ (2005, p. 30) – unaffected, in other words, by the scholar’s confined subjective outlook.

Quite unsurprisingly, such statements have led to fierce criticism. In a review of Moretti’s book *Distant Reading* (2014), Shawna Ross argues that

when he claims that his grandiose stories are ‘resting solidly on facts’ [Moretti 2013, 44] or that distant reading yields ‘the clarity of the empirical confirmation,’ [Moretti 2013, 92] the generalization-spouting bravado by which Moretti skates over impossibly broad terrain finally comes across as glib, revealing one of the primary dangers of digital literary studies to be the adoption of an aggrandized, even hubristic attitude toward literature as so much inert stuff being poked at. (Ross, 2014, p. 5)

Some of the common arguments against distant reading are used by Ross: the positivism in such allegedly facts-based research, the naive belief in generalizations over ‘impossibly broad terrain’, and the arrogance – ‘aggrandized, even hubristic attitude’ – implicit in all this ‘bravado’. Others, such as computational critic Stephen Ramsay in *Reading Machines* (2011) and philosopher Tom Eyers in *Speculative Formalism* (2017), have also warned against the rise of such neopositivist premises in literary studies.

Two general points of criticism against the digital humanities are relevant here. First, a problem-solving rhetoric might overshadow the interpretation and theorization of the findings of data-driven literary studies in the spirit of Jockers and Moretti. When the main issue becomes how to solve a methodological problem (e.g., how to automatically detect characters in literary texts), the theoretical motivations behind such methodological problems tend to become
background noise (Scheinfeldt, 2010). Second, a building-epistemology rhetoric likewise obscures how data analysis and digital technology might contribute to theoretical issues. According to Natalia Cecire, the widespread focus on the building of algorithms, tools, models, and databases invokes the impression that making technical constructs is the primary aim of data-driven humanities research, which she believes conforms to a neoliberal logic (2011). In the field of Dutch literary studies, these points of criticism are largely reflected in an article by Stephan Besser and Thomas Vaessens on the emergence of the digital humanities in the Netherlands, warning against its alleged scientificity, its uncritical essentialism, and its lack of reflection on theoretical assumptions (2013).

Scientific ideals such as replicability and generalizability as propagated under the headers of distant reading (Moretti) and macro analysis (Jockers) are not, however, new to literary studies. For the Dutch situation, the influential literary journal Merlyn (1962–1966), with which the work of J. J. Oversteegen and A. L. Sötemann is frequently associated, embodied the ideal of replicability. In line with structuralist theory, the journal envisioned a fundamental text-centric approach to literature. In their first issue, the editors explicitly mention a ‘criterion of replicability’ literary analysis should conform to. The ideal of generalizability is explicitly put forward by Suzanne Fagel in her work in the field of Dutch literary stylistics (Fagel, 2015). Arguing for a quantitative approach to literary style, she contends that ‘an “introspective” method does not provide a reliable basis for judgements about how often something appears in a text, or how generally shared (generalizable) a certain interpretation is’ (Fagel, Stukker, & van Andel, 2012, pp. 180–181).

More generally, the oppositions between close and distant reading, micro and macro analysis, computer-oriented and philologic study, as invoked in these debates, are highly schematic. Although Jockers (2013) and Moretti (2013) adopt a polemical attitude toward ‘traditional’ forms of literary study, they do not argue for a replacement of micro analysis and close reading but opt for a blended, mixed approach. In fact, it would be hard to find a scholar using digital techniques stating that the human, close, micro element has become obsolete. In the Netherlands, this is exemplified by statements of three professors of literary studies and/or digital humanities. In his inaugural speech, professor of digital humanities Rens Bod proposes that the computational search for patterns (distant) should be combined with critical reflections on those patterns (close) (2013). In her inaugural speech, professor of computational literary studies Karina van Dalen-Oskam emphasizes that both the quantitative, measurable component (distant) and the cultural component (close) of literary phenomena
should be taken into account (2012). In an article on the gap between close and distant reading, professor of early modern literature Els Stronks contends that both approaches should complement, broaden, and nuance one another (Stronks, 2013, p. 213).

In line with the pleas of Bod, Van Dalen-Oskam, and Stronks for a productive dialogue between methodologies, Paul Fleming (2017) observes that a recent strand of data-driven literary studies has emerged that goes beyond the schematic close-distant reading opposition. Starting from the observation that ‘close reading is always exemplary in a double sense: the exemplary reading of exemplary passages’ (p. 437), he argues for the importance of a ‘recursive relation’ between exemplary close reading and computational modelling. After disseminating the fundamental opposition between close and distant reading in Moretti’s work (with Moretti as the classic example scholar of distant reading), he encourages us ‘to look to other examples in and of digital humanities’ (p. 453). These other examples can be found in the ‘new wave of scholarship in the digital humanities’, exemplified by scholars such as Andrew Piper, Hoyt Long, Richard Jean So, Alan Liu, and Ted Underwood, who ‘[insist] upon bringing together close and distant reading, computational analysis and exemplary exegesis as inextricable from one another’ (p. 439).

The post-Moretti era of distant reading is characterized by a strong emphasis on transcending narrowly defined boundaries of close and distant reading, as well as by a general focus on recursive modelling. Recursively going back and forth between distant and close, macro and micro, data and interpretation, numbers and words, these scholars showcase the importance of creating multidimensional analytical models of literary texts. Subsequently, this type of research does not easily fall prey to the neopositivist assumptions outlined above. This anti-positivist stance is best exemplified by the work of Stephan Ramsay. In Reading Machines (2011), he states that ‘the scientist is right to say that the plural of anecdote is not data, but in literary criticism an abundance of anecdote is precisely what allows discussion and debate to move forward’ (p. 9). Contrary to Jockers’s pejorative reference to ‘anecdotal evidence’ (2013, p. 5) as insufficient means, Ramsay stresses that literary studies, he contends, is not meant to settle these discussions or to solve problems emerging from them. Conversely, Ramsay emphasizes that ‘literary criticism operates within a hermeneutical framework in which the specifically scientific meaning of fact, metric, verification, and evidence simply do not apply’ (2011, p. 7), which is in clear opposition to Moretti’s claim that computation has the potential to ‘falsify existing theoretical explanations’ (2005, p. 30).
Following Ramsay, this book does not aim to settle debates on the representation of social groups in present-day Dutch literature once and for all but hopes to open up empirically informed discussions related to the topic. By building, applying, and evaluating recursive models of representation, it does attempt to formulate data-driven generalizations on the topic without denying that these generalizations are the result of specific choices made in the shaping of these models. As explained in the next subsection, the opportunities and dangers of modeling are a central issue within the research associated with cultural analytics.

1.3.2 Modeling in Cultural Analytics

The data-driven critique of representation presented in this book is closely aligned with studies carried out under the header of cultural analytics, a subfield of the digital humanities that studies culture through computation. While originally coined by Lev Manovich in 2005, the term ‘cultural analytics’ is currently first and foremost associated with the peer-reviewed, open-access *Journal of Cultural Analytics* (2016–present), edited by Andrew Piper. Manovich and Piper opt for relatively similar definitions, but whereas Manovich’s definition is slightly more narrowly focused on media theory and visual data analysis, Piper’s use of the term encompasses a broader terrain. In the journal’s opening article, Piper states that cultural analytics is more than ‘computer science applied to culture’ as it ‘requires a wholesale rethinking of both of these categories’ (2016):

> Computation forces us to rethink our current disciplinary practices in the humanities from the ground up. What counts as evidence? What is the relationship between theory and practice? How do we account for the technological mediations of our critique? But culture too impinges upon computation. It challenges the universalism and the neutrality implicit in many computational applications. It reminds us that knowledge is always situated, somewhere, at some time, by someone. Putting culture into computation cautions us to remember where we are when we think we know something. (Piper, 2016)

The two-way street between culture and computation as proposed by Piper contains a fundamental recursivity prompting scholars to go back and forth between the practices of cultural study and those of computational analysis. Using computation to study culture forces a rethinking of what David Berry calls the ‘hard core’ of the humanities (2011, p. 4), for which it is required to
make explicit the implicit assumptions about, for instance, the role of evidence and the dynamics between theory and practice. Conversely, cultural study incites computational study to be conscious of its situatedness, its semantics, and its biases and provides it with critical theoretical frameworks. Just as the analysis of cultural products cannot make any claims to neutrality and universalism, neither can writing a set of instructions in the form of an algorithm. Behind both cultural and computational analysis are selective choices that emerged from the perspective adopted.

Such cross-fertilizations between culture and computation highlight the importance of modeling. What, then, is modeling? In *Distant Horizons* (2019), Ted Underwood gives a most straightforward definition of a model as ‘a relationship between variables’ (p. xii). Reversing the steps of inquiry as proposed by Moretti (2017), Underwood furthermore emphasizes the importance of hypothesis testing: ‘Instead of measuring things, finding patterns, and then finally asking what they mean, we need to start with an interpretive hypothesis (a “meaning” to investigate) and invent a way to test it’ (2019, p. 17). A simple example: hypothesizing that female authors write longer sentences, the variables ‘sentence length’ and ‘author gender’ can be inserted in a model to test whether an author’s gender is predictive of the length of their sentences. Applying this model to a corpus of both male- and female-authored texts might result in the claim that female authors indeed write longer sentences, which then should be further tested and interpreted. More metaphorically than Underwood, Piper defines a model as ‘a metonymical tool – a miniature that represents a larger whole’ (2016). In the example on sentence length and author’s gender, the ‘miniature’ model is applied to a corpus taken to be representative of the ‘larger whole’ of sentence length in, say, nineteenth-century literature. Importantly, Piper stresses the fact that a model is ‘also recursive in that it can be modified in relationship to its “fit,” how well it represents this whole’ (ibid). Based on the model’s performance, there might be reasons to think that the author’s gender is not the best predictor for sentence length, in which case the model is a rather poor representation of the ‘whole’ of sentence length in nineteenth-century literature. The model can then be modified again and again – perhaps other variables such as an author’s age or author education should be considered – until there are solid reasons to claim that the model ‘fits’, within certain statistical bounds of likelihood, the whole it attempts to represent.

Apart from the methodological ‘benefits of speed, automation, and scale that computational representations afford’ (Ramsay, 2011, p. 8), the importance of modeling shows that there are sound theoretical reasons for using computation.
to study culture. Modeling does not only provide scholars with 'a science of generalization' (Piper, 2018, p. 9), it also prompts them to think about ‘the constructedness of knowledge and the observer’s place within it’ (ibid). As this book studies the ways in which social groups are represented in Dutch literature, it is, furthermore, relevant that models are ‘first and foremost representations, miniatures that mediate between ourselves and our observations’ (ibid., emphasis added). Mediation not only takes place within processes of literary representation (see section 1.2.1), it also has a vital function within the study of these representations as presented in this book. Going back and forth between numbers and words, it aims to make explicit the situatedness of the knowledge resulting from the constructed and reconstructed models presented in the following chapters.

Studies in cultural analytics published in Journal of Cultural Analytics regularly take their cue from socially indebted or ideologically oriented perspectives, which is exemplified, among others, by a range of data-driven analyses of gender and race representations in literature. Such topics lend themselves perfectly to recursive modeling: as cultural representations of social reality are always communicated by a human mediator (see section 1.2.1), it seems only fair to make the role of mediation explicit in the study of these representations. As the role of mediation often remains implicit in close reading–based critiques of representation, it tends to be unclear how the findings came about. This is particularly salient in light of Stephan Ramsay’s observation that ‘the critic who endeavors to put forth a “reading,” puts forth not the text, but a new text in which the data has been paraphrased, elaborated, selected, truncated, and transduced’ (2011, p. 16). Each – close or distant – reading of a text transforms the original text into something else by selecting parts of it, paraphrasing those parts, elaborating on them, zooming in or zooming out, and so forth. These selections, paraphrases, elaborations, truncations, and transductions are all mediations considered by the scholar to be representative of the original text. Recursive modeling as practiced in this book is an attempt to make these mediating steps explicit.

Within the discussion on symptomatic reading versus surface reading as kick-started by Stephan Best and Sharon Marcus (2009), the recursive models of literary representation as carried out in this book provide an argument for a data-driven critique of representation. With the term ‘surface reading’, Best and Marcus presented an alternative to the dominant, ‘symptomatic’ form of reading in academia, a mode of reading in which textual elements are considered as symbolic or symptomatic of a deeper lying societal issue.
Whereas symptomatic readings aim to uncover the latent, invisible, silent, or repressed meanings of texts, a surface reading focuses on what is evident, perceptible, apprehensible in texts; what is neither hidden nor hiding; what, in the geometrical sense, has length and breadth but no thickness, and therefore covers no depth. A surface is what insists on being looked at rather than what we must train ourselves to see through.’ (Best & Marcus, 2009, p. 9)

Critiques of representation in Dutch literature in the tradition of *De canon onder vuur* (see section 1.2.2) are often symptomatic in the sense that they attempt to read between the lines and in the margins of the text, in order to lay bare hidden, and sometimes inconvenient truths about the work under scrutiny. Of course, such analyses do also consider what happens on the surface of the text (e.g., which words are used to characterize a specific character?), but they tend to center on what happens below that surface (e.g., which characterizations of a specific character remain implicit?). Conversely, surface reading is proposed by Best and Marcus as a means to center on what the text says instead of what it does not say. Quite unfortunately, however, Best and Marcus fall prey to the same kind of neopositivism as Jockers and Moretti (see section 1.3.1) when they state that ‘digital modes of reading may be the inspiration for the hope that we could bypass the selectivity and evaluative energy that have been considered the hallmarks of good criticism, in order to attain what has almost become taboo in literary studies: objectivity, validity, truth’ (p. 17). Despite the flawed, naive positivism ascribed to surface reading by Best and Marcus, their notion of the textual surface does serve as a convenient metaphor to illustrate the relevance of a data-driven, as opposed to a solely qualitative, critique of literary representation. As Terry Eagleton states in *Ideology* (1991): ‘To study an ideological formation, then, is among other things to examine the complex set of linkages or mediations between its most articulate and least articulate levels’ (p. 50). The ideological formations present in the representations of social groups studied in this book can be broken down in an articulate level (the surface of the text) and a less articulate level (the hidden depth of the text). Recursive modeling is a means to make the ‘linkages or mediations’ between these two levels explicit. Inserting computation into the study of these representations thus helps to show which textual elements at the surface of the text might lead to statements on the depths of the text.

Best and Marcus do not fail to notice that ‘to see more clearly does not require that we plumb hidden depths and that producing accurate accounts of surfaces is not antithetical to critique’ (2009, p. 18). The models presented in
the following chapters are examples of how this might work in practice. In the
data-driven components of these models, hypotheses on the representations of
social groups are first tested by ‘producing accurate accounts of [the] surfaces’
of each of the 170 texts in the corpus. Statistical analysis is used to either reject
or confirm the hypotheses. But although these statistical findings are based on
what happens on the surfaces of the texts, they might also give a clue as to what
happens ‘underneath’ them. In chapter 3, for instance, the hypothesis is tested that
male and migrant characters have lower centrality scores in their co-occurrence
networks than female and nonmigrant characters. A predictive statistical model
then attempts to arrive at the probability that this is the case solely based on the
coop-occurrences of characters visible on the textual surfaces. The outcome of this
model already suggests something about the deeper, hidden meanings related to
representation of gender and descent, specifically regarding the dominance of a
type of character. But the centrality of characters is, perhaps quite obviously, not
fully covered by such a statistical, surface analysis: stylistic, thematic, and other
dimensions also contribute to what makes a character important in a narrative.
Qualitative assessment – close reading – is then used to evaluate the power of
these statistical patterns for specific cases. Whereas the statistical model focuses
on textual repetitions at the surface of the novels, the close readings attempt
to uncover textual rarity that might possibly provide an argument about how
characters in a specific text are central in other than statistical terms.  

1.4 METHODOLOGICAL BACKGROUND

1.4.1 Tools: Narratology and Network Analysis

The twofold theoretical framework as described in the previous sections is
operationalized in the next chapters by using tools often used in each of the two
theoretical strands. In critiques of literary representation, the methodological
toolkit of narratology aids in coordinating the qualitative assessment of texts.
Studies in cultural analytics can rely on a wide variety of computational or
statistical tools, but social network analysis provides the most convenient tool for
the study of the relational patterns within the fictional populations of characters
in the corpus.
While the term ‘narratology’ was coined in 1969 by the Bulgarian linguist Tzvetan Todorov, its theoretical origins can be traced back to the early years of formalism at the beginning of the twentieth century (Meister, 2011). Narratology focuses on formal characteristics of texts that are constitutive of the (overall) narrative structure(s). In narratological analysis, interpretations of texts are centered on more or less delineated narratological concepts. One of the concepts that is used throughout each of the following chapters is characterization, as the study of the representation of social groups almost always requires insights into the ways a character is characterized. The narratological concepts of narration and focalization play a central part in chapter 3, ‘Centrality’. An important part of the centrality of characters belonging to a social group is the ways in which they (are) narrate(d), as well as how they perceive the narrative world and how they are perceived by other characters. In order to gain insight into the one- or multi-voicedness of the groups in which characters function, chapter 4, ‘Community’, primarily uses the concept of polyphony as defined by Mikhail Bakhtin. Chapter 5, ‘Conflict’, makes use of the actantial model as described by A. J. Greimas to grasp the hierarchies in conflicts between characters belonging to a specific social group.

Social network analysis has a range of convenient metrics to dissect the relational structures between characters. A wide range of such metrics is contained in the software library Networkx written for the Python programming language used in most data-driven analyses in this book. Before applying those metrics, it is necessary to create network representations of the fictional populations in the corpus. Chapter 3 first describes a pipeline for extracting social networks of characters from each of the 170 novels, which will be the basis of the data-driven components of each of the chapters. In order to calculate the importance of social groups in statistical terms, this chapter then scrutinizes the social networks through a range of centrality metrics. The one- or multi-voicedness of character communities are studied in chapter 4 by breaking down the 170 networks into subnetworks of characters with a community-detection algorithm, as well as by computing so-called homophily, i.e., the extent to which any two characters share similar features in the networks. Chapter 5 models the dominance or subordination of character types in conflict situations by calculating a ‘conflict score’ for each character and then more generally models the extent of antagonism or social balance in the networks by analyzing triangular configurations of friends and enemies.
1.4.2 Corpus and Data

The corpus consists of all 170 submissions to one year of the Libris Literatuur Prijs (hereafter: the Libris prize), one of the most prestigious literary prizes in the Dutch language area. As all new novels written in the Dutch language are eligible for the prize, the list of submissions consists of both Dutch and Flemish authors (more on the background of the authors in chapter 2). In this book, ‘Dutch literature’ thus refers to both Dutch and Flemish literature, as is common in the Dutch language area. The year 2013 is a randomly chosen sample year of the contemporary production of Dutch literary fiction. As the Libris prize targets novels published in the previous year, all novels in the corpus were published in 2012. The prize roughly follows the system used by the Man Booker Prize for Fiction. There is no restriction on the number of novels publishers can annually submit. From the aggregated list of all submitted novels, members of an annually changing professional jury first select a longlist of 18 titles, and then a shortlist of 6 titles. As publishers submit novels that they hope to be possible winners, the full list of submissions reflects what they see as high-quality literary fiction. Since 2010 the prize targets literary novels for adults exclusively. The consequence is that other forms of prose cannot be submitted: the bulk list of submitted novels does not contain young adult novels, children’s books, fantasy, or (literary) thrillers. As genre boundaries are obviously fluid, some Libris submissions could, however, to a greater or lesser extent be characterized as something other than literary fiction. Like the BookSpot Literatuurprijs (earlier known as the ECI Literatuurprijs, AKO Literatuurprijs, and Generale Bank Literatuurprijs) and the Fintro Literatuurprijs (earlier known as Gouden Uil), the Libris prize grants a relatively high amount of money to laureates. Each of the six shortlisted authors receives €2500. On top of that, the winner of the prize receives €50.000. Winning or being on the shortlist of the Libris prize thus not only leads to an increase in symbolic capital, it also affects the author’s economic capital.

The corpus of 170 novels constitutes a fair share of the production of literary fiction in the sample year 2012. According to the database of the KB, the national library of the Netherlands, 1,475 books with NUR-code 301 (literary fiction) were published in 2012. Subtracting duplicates, reissues, and exclusively online publications from this list results in a total number of 460 works of literary fiction originally published in Dutch. The sample of 170 novels thus represents 36.9 percent of the Dutch books of literary fiction published in this year. Compared to sample sizes in sociological research, the sample size of the current research
is highly representative of the ‘population’ of books published in this genre in this year.⁶³

In light of most qualitative, close-reading based scholarship, 170 novels is a rather large sample, but it is relatively small in comparison to other studies in the field of cultural analytics. This is especially true for studies focusing on English language texts, which can rely on extensive databases of texts such as the Hathi-Trust Digital Library, which is for instance used by Ted Underwood in *Distant Horizons* (2019). Andrew Piper uses a corpus of 7,500 English language novels containing 650,000 characters for the analyses presented in his chapter on characterization in *Enumerations* (2018), which is more than 44 times the size of the sample used in this book. The major benefit of the current sample size is that it makes manual annotation of character features possible. Up to the present, it is only possible to automatically detect the gender of characters by using programming pipelines such as BookNLP (Bamman et al., 2014; Piper, 2018; Underwood et al., 2018),⁶⁴ but this is done on the basis of predictions and is thus not as accurate as manual annotation.⁶⁵ Demographic features such as country of descent, level of education, age, and professional occupation are probably never fully automatically detectable, as they often remain implicit in the text and thus require a fair amount of interpretation. Furthermore, a sample size of 170 also enables a more efficient qualitative evaluation of algorithmic and statistical analyses than would be the case with a larger sample size. It requires less effort to check for, say, 10 percent of the 170 novels if the output of the algorithms match up human intuitions than it would be to do this for 10 percent or even 1 percent of 7,500 novels.

The analyses in the following chapters are based on an extensive collection of metadata – gender, country and place of descent, country and place of residence, age, level of education, profession, or daily activity – on 2,137 characters, as identified by human annotators. The data collection was carried out in several phases roughly between 2014 and 2018.⁶⁶ In 2014 a group of annotators contributed to the first phase of data collection, resulting in metadata on 1,176 characters.⁶⁷ As the guidelines for annotation were not as precisely defined in the first iteration, the annotation process was repeated twice with more clearly defined guidelines. Most importantly, a formal threshold was introduced to be able to distinguish which entities in a text should or should not be adopted in the database as characters.⁶⁸ In the period from 2017 to 2018, two student assistants increased the number of characters to a total of 2,137 characters and complemented some of the missing metadata in the earlier database.⁶⁹ Furthermore, these research assistants added relational information between characters (when known) –
friend, enemy, colleague, lover, family – to the database. All collected data and the software created for this book are available via an open-access GitHub repository, \(^70\) which, however, does not contain the digital versions of the 170 texts of the novels due to copyright limitations. \(^71\)

A deliberate reductionism is inherent to the process of data collection. First, not all demographic features are deducible from the narrative worlds. Contrary to most sociological survey data collections, a considerable number of features is unknown for the 2,137 identified characters in the corpus. \(^72\) Second, literary texts have the possibility to (de)construct various definitions of, for instance, gender, as an author can play with notions of femininity and masculinity. In most cases, the annotations of the demographic categories cannot account for such artistic strategies. Third, the annotations are inevitably binary: this either/or logic does obviously not cover the variety of identities within one demographic category. This is not to say that the fluidity of these categories are outside the scope of this book; in the close reading of the case studies, attention is devoted to the ways in which such seemingly fixed categories are commented on.

1.4.3 Previous Research on Corpus and Dataset

The whole and parts of this dataset have been used in studies published earlier. Initially, a study on the first data collected in the period 2014–2015 was published in *Journal of Dutch Literature* in 2016, authored by Lucas van der Deijl, Saskia Pieterse, Marion Prinse, and myself. This article describes and interprets the diversity within the demographic landscape in the novels demonstrating that male, Western, higher educated characters are the dominant voices in the present-day Dutch literature the corpus is considered to be representative of. Although this article was granted the annual academic prize 2018 for best article published in 2016 and 2017 by the Society of Dutch literature (Maatschappij der Nederlandse Letterkunde), it was also met with fierce criticism. \(^73\)

Together with professor of sociology Beate Volker, I coauthored the article ‘Imagined Social Structures: Mirrors or Alternatives? A Comparison between Networks of Characters in Contemporary Dutch Literature and Networks of the Population in the Netherlands’ in *Poetics* (2019). Based on 1,397 characters of the dataset (the second phase of data collection), this article compares the networks of characters in present-day Dutch literature with the actual networks of the present Dutch population. It argues, among other things, that social networks in Dutch literary fiction are less segregated in terms of descent than actual social networks. Although this comparative line of research between literature and
society is not elaborated upon in the remainder of this book, it is a promising perspective on the million-dollar question as to how fiction reflects society. This book is indebted to elements of the theoretical and methodological frameworks of these preliminary studies, and the subsequent chapters build on their findings.

1.5 STRUCTURE OF THE BOOK AND INSTRUCTION FOR READING

The second chapter, ‘Data’, contains an overview of the data on which the analyses presented in the third, fourth, and fifth chapters are based. It describes some basic statistics of the 2,137 characters populating the 170 novels in the corpus. In chapters 3, 4, and 5, an answer to each of the sub-questions is formulated (see 1.1.2 for an elaboration on the research questions). These chapters are structured in a similar fashion. Following the twofold theoretical framework and the mixed-methods setup presented in this introductory chapter, each of these three chapters starts with an elaboration of how its central concept (centrality, community, conflict) is theorized and operationalized in both network theory (a method used in data-driven research in cultural analytics) and narratology (a method used in qualitative critiques of literary representation). Based on these insights, one or more data-driven models are described and hypotheses are formulated. After the statistical patterns resulting from these models are described, their significance is assessed through one or more close readings of case studies from the corpus. Combining quantitative, network analytical, statistics-based analysis with qualitative, narratological, close-reading based analysis, these chapters demonstrate how centrality, community, and conflict of characters affect the representation of social groups in the corpus. Chapter 3 first describes the book’s approach to extract fictional social networks from the texts in the corpus and then presents a model to predict which types of characters are central in terms of network structure. The results are then evaluated in light of a close reading of how certain social groups are depicted as dominant or subordinate in Özcan Akyol’s Eus (2012).

Chapter 4 presents two models to detect communities of characters. The first model uses community detection algorithms to break down each of the 170 novels into distinct groups in order to test the extent of integration or segregation between genders, descents, classes, and ages. The depiction of communities in
Philip Huff’s *Niemand in de stad* (2012) is then read against the background of the pattern generated by the model. In the second model, so-called ‘homophily’ is computed between every two characters in order to assess how characters of a certain gender, descent, education, and age tend to flock together in the narratives, the results of which are then evaluated through a close reading of how youth, old age, and death are represented in Mensje van Keulen’s *Liefde heeft geen hersens* (2012).

Chapter 5 proposes two models of narrative conflict. Based on one-on-one conflicts between characters, the first model is used to test which types of characters are the dominant parties in conflicts between two characters. The resulting statistical pattern is then assessed by narratologically breaking down the conflict between classes in Bart Koubaa’s *De Brooklynclub* (2012). The second model tests the extent of social balance in conflicts between three characters. Leon de Winter’s *VSV* (2012) and Tommy Wieringa’s *Dit zijn de namen* (2012) are then used to qualitatively demonstrate how social balance in such triangular conflicts to a greater or lesser extent results in a (schematic) moral opposition between social groups.

In order to answer the general research question ‘How are social groups represented in present-day Dutch literary fiction?’, the concluding, sixth chapter brings together the findings of the chapters on centrality, community, and conflict. It furthermore demonstrates the book’s theoretical and methodological contribution to the field of cultural analytics and character-based critiques of representation. Finally, it evaluates the study’s limitations and strengths and proposes directions for further research on the topic.

Because of the book’s interdisciplinary approach, it targets a double audience. Its aim is to interest both scholars working in the qualitative strands of literary and cultural criticism and those experienced in quantitative, statistical, or computational methods. As each of these strands of research has its own academic style and conventions, it is a challenge to conform to prevalent norms in both strands, which I, however, have attempted in this book.

In order to avoid a confusion of tongues between audiences, I have clarified where necessary common narratological concepts and terms as well as the workings of statistical tests and computational techniques. It is, of course, very possible that readers not familiar with these concepts, terms, or techniques still have a hard time following along. If that is the case, there is always the possibility to skip the more technical parts of the book (and take my word for it) and start reading after the statistical results are reported. Likewise, readers who are primarily interested in the computational techniques can of course also focus
more on these than on the literary theorizations and the close readings of case studies. Both types of readers can turn to the conclusion to each chapter for the main findings.