

THE RIDDLE
OF
LITERARY
QUALITY



A COMPUTATIONAL APPROACH

**A
U
P**

KARINA VAN DALEN-OSKAM

The Riddle of Literary Quality

The Riddle of Literary Quality

A Computational Approach

Karina van Dalen-Oskam

Amsterdam University Press

Originally published as: K. van Dalen-Oskam, *Het raadsel literatuur: is literaire kwaliteit meetbaar?*
Amsterdam University Press, 2021

Translated and adapted from the Dutch by the author.

Cover design: Margreet van de Burgt

Lay-out: Crius Group, Hulshout

ISBN 978 90 4855 814 8

e-ISBN 978 90 4855 815 5

DOI 10.5117/9789048558148

NUR 621



Creative Commons License CC BY NC ND

(<http://creativecommons.org/licenses/by-nc-nd/4.0>)

© K. van Dalen-Oskam / Amsterdam University Press B.V., Amsterdam 2023

Some rights reserved. Without limiting the rights under copyright reserved above, any part of this book may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise).

Contents

1. The Riddle	9
<i>Introduction</i>	
The Riddle of Literary Quality	9
From Sociology of Art to Neo-Darwinism	13
Computational Literary Studies	18
Empirical Study of Literature	27
Conclusion	28
References	29
2. The National Reader Survey	32
<i>Books, Readers, Opinions</i>	
The Survey	32
The Books	35
The Readers	37
Categories of Books	42
Winners and Losers	44
Motivations	51
Contrasting Opinions	55
Conclusion	60
References	61
3. Romance, Suspense, and Translations	63
<i>Mixed Methods in More Detail</i>	
From Readers' Opinions to Text Analysis	63
Romance and the <i>Fifty Shades</i> Trilogy	66
Crossing the Borders of Suspense	77
The Riddle of Translation Quality	80
Translations in the Periphery	89
Conclusion	91
References	92
4. Literary Novels Written by Women	95
<i>Donoghue and Dorrestein</i>	
The Glass Ceiling	95
Translations Compared	100

Donoghue, <i>Kamer (Room)</i>	103
The Dutch Perspective	106
Dorrestein in More Detail	115
Dutch Originals Opposed	121
Conclusion	122
References	123
5. Literary Novels Written by Men	124
<i>From Barnes to Grunberg</i>	
Barnes and the Booker	124
Translations Opposed	127
Dutch Originals Compared	132
A Literary Duel: Kluun versus Koch	141
Grunberg and Romance	148
Conclusion	154
References	156
6. Style, Gender, and Genre	158
<i>Combining Multiple Perspectives</i>	
Beyond Words	158
Syntactical Analysis	159
Parts of Speech	160
Cliché Expressions	161
Themes or Topics	162
Predicting Scores for Literary Quality	163
Vector Space Explorations	166
Sentence Length	168
Comparing Dutch Literary Novels by Female and Male Authors	170
Author Gender and Perceptions of Literary Quality	177
The Feedback Loop of Genre	181
Conclusion	184
References	186
7. The Riddle of Literary Quality Solved?	188
<i>Reflection and Conclusion</i>	
References	198

Appendix 1: The Survey	200
Appendix 2: The Books	204
Appendix 3: The Website	214
Bibliography	215
Tables	222
Figures	223
Acknowledgements	225
Index	227

1. The Riddle

Introduction

Abstract: This chapter introduces the project The Riddle of Literary Quality in which computational text analysis methods were applied to help understand what makes contemporary novels literary or not in the eyes of their readers. The chapter concisely describes the institutional approach of Bourdieu's sociology of art, the new sociology of art in which the artwork itself is placed back into the picture, and neo-Darwinian thoughts about art. Next follows a short overview of the affordances of computational literary studies and of the empirical study of literature. All disciplines together form the background against which The Riddle of Literary Quality took a totally new, combined approach to research into literary value attribution.

Keywords: Literary quality, Bourdieu, Sociology of Art, Empirical Study of Literature, Computational Literary Studies, Distant Reading

The Riddle of Literary Quality

This book gives an overview of the results of the Dutch research project The Riddle of Literary Quality ('The Riddle') and presents the information and data needed to verify and replicate most of the research. The aim of the project was to apply computational methods to answer the question whether we can somehow measure the literary quality of novels in the texts themselves. Which linguistic features correlate with readers' perception of literary quality? The project was funded by the Royal Netherlands Academy of Arts and Sciences in its Computational Humanities Programme and ran from 2012 until 2019.

The quest to uncover conventions of literariness using computational means was deliberately limited to the Netherlands, the small country in Western Europe where I live and where I was born and raised. In 2013, the Netherlands had a population of just under 17 million. The official language of the country is Dutch, a West Germanic language in the family of Indo-European languages. The Dutch language is also

native to Flanders, a part of Belgium, the country bordering the Netherlands in the South. The literary landscapes of the Netherlands and Flanders partly overlap but they each have their own characteristics, which informed the limitation to a country and not to a language area for this research. In the ‘world republic of letters’, the Netherlands and Belgium both figure in the periphery. The centre to which this periphery mainly belongs is the English language and culture of predominantly the United States and the United Kingdom. As is usual for language areas and countries in the periphery, publishing houses and bookstores in the Netherlands provide readers with a large amount of translations. In 2009, translations from other languages into Dutch in the categories of fiction and poetry amounted to 65 to 70 per cent (Bevens et al. 16–17, 386; Franssen 30–37). This abundance of translations on the Dutch market is a fact not to be ignored. The Riddle team chose to analyse not only works written and published in Dutch but to also include novels translated into Dutch; a sound introduction to Dutch literature in translation is *Doing Double Dutch. The International Circulation of Literature from the Low Countries* edited by Brems et al.

In this monograph I will carefully describe how the research was organized, which methods were used, what could be observed in the results, and to which next steps these observations may lead. Because of this focus on a synthesizing overview, the references I add are mainly publications that directly influenced the project. From the many more recent publications that may also be relevant, I mention only a few that may help those interested in the field along the way. For more detailed information the interested reader also might want to turn to the two PhD-theses and to the other papers that the project delivered. The same goes for more information about the data, the selected methods, and details about the statistics that came out of the project.

The selected computational approach and its affordances go hand in hand with a very cautious attitude towards over-interpretation of the outcomes. Instead of elaborately evaluating the results from the perspective of earlier work in traditional literary theory and drawing firm conclusions, I will repeatedly emphasize that although the results do significantly increase our knowledge about the conventions of literary quality in the Netherlands around 2013, there remains a lot to be learned. Bigger corpora, more data from readers, and more mixed-methods research combining non-traditional and traditional methods are needed to confirm many of the observations and deepen them with additional explorations.

In the Dutch academic field of literary scholarship the work of the French sociologist Pierre Bourdieu has long played and still plays a very important role. Bourdieu’s ‘field theory’ explains how value is attributed to works of art in an intricate set of sociological processes that do not have much to do with the content of the art object but primarily rest on connotations with prestige and class (more

on that in the next section of this chapter). However, anecdotal evidence abounds that this sociological explanation of what art is, is highly counter-intuitive to most readers, professional or not. For them, it is extremely unsatisfactory that literary scholars cannot explain why a certain novel is regarded as highly literary and why another novel is seen as 'trash'. Many readers would like to get concrete pointers to what contributes to the actual literary quality of the texts themselves. They assume literary quality is a feature that can be established objectively. In a still famous talk given in Leiden in 1978 and published in 1979, the then popular Dutch author, critic, and scholar Karel van het Reve criticized literary scholars for their unreadable scholarly work and for ignoring what he posited is the key question in their scholarly discipline:

One of the key questions of literary theory should really be: Is it possible to make a description of a good book, without using the word 'good', which is not applicable to a worthless book? And vice versa: Is it possible to describe a bad book, again without using the word 'bad' or similar words, in such a way that it is impossible to make or find a good book that fully fits that description? This crucial and interesting question is avoided like the plague in literary theory. This question has been eliminated in one fell swoop by placing poor-quality literature outside of literature, by excluding from research one of the most interesting phenomena of literature, difference in quality. (Van het Reve 135–136, my translation; cf. Van Dalen-Oskam (2021) 18–19)

Present-day Dutch literary scholars still feel uneasy with this lecture and even now avoid answering the question Van het Reve posed in such daring words. To this day, informed readers – old and young – regularly refer to his talk when they talk about literature. And literary quality still is a hot topic in the Netherlands. In general, Dutch readers are aware of the literary quality of what they read, think they should read, or avoid reading. The project The Riddle of Literary Quality aimed to find out what the differences are between novels perceived by Dutch-language readers as highly literary and those seen as having hardly any literary quality. To this end it brought together the results of two approaches: first, a large survey of readers' opinions about four hundred novels (called *Het Nationale Lezersonderzoek – The National Reader Survey*) and second, an analysis of the same four hundred novels using methods of computational text analysis. Questions the project set out to answer were for example: Which features in novels may have played a role in Dutch-language readers' perceptions of literariness in 2013? What were the conventions of literary quality in the Netherlands at the time of our survey? What does this tell us about Dutch society, in which these conventions are embedded, and what does it suggest about the past? To which new research questions do these results give rise?

Apart from providing an answer to these questions in this monograph, it is my specific aim to present such a thorough description of the research process and of its results that the research can be replicated in other countries or language areas. The Dutch situation thus will function as a point of reference for other projects and may help to find similarities and differences in the conceptions of literariness and the functions of literature in other cultural and geographical areas. At the time of writing this book, I was involved in the very first instalment of such a replication in the United Kingdom. At the University of Wolverhampton, Professor Sebastian Groes and I led a project funded by the Arts and Humanities Research Council in which we gathered readers' opinions about four hundred novels through a large survey that we named the Big Book Review. As in the Dutch project, we will correlate the outcomes of this survey with the results of a computational text analysis of the selected novels. The results of the project in the United Kingdom, called *Novel Perceptions: Towards an Inclusive Canon*, will be presented in book form in due time.

Making research replicable includes more than a clear description of what has been done. It also requires that project data are made available. For the Dutch project, this was done both in making data available in csv files and in an R package that can be installed from CRAN (The Comprehensive R Archive Network) and can be run in R. Detailed information about the 'litRiddle' package is available in the package itself and on the website accompanying this book. A link to the website can be found in Appendix 3.

I designed and led *The Riddle of Literary Quality* and conducted the project together with a whole team of researchers, developers, and students. Two English-language PhD-theses resulted from the project, by Andreas van Cranenburgh (2016) and Corina Koolen (2018). Together with other team members we authored several peer-reviewed articles about the work done in the project. Most results of my own research are published for the first time in English in this book; I presented them in a slightly different form in a more elaborate Dutch book about the project that I published with Amsterdam University Press in 2021. Whenever the research presented is not exclusively done by me, I will always explicitly mention who else contributed to the work.

The remainder of this introductory chapter will first delve deeper into the sociology of art and related work. Then it will describe how *The Riddle of Literary Quality* positions itself in the booming research discipline of computational literary studies and address aspects of earlier research in this tradition on which it builds. I will also position the research in relation to empirical literary studies. Chapter 2, *The National Reader Survey*, carefully explains the organization of the survey and the selection of the corpus. It also presents the general results of the survey and how these informed the organization of the next chapters of the book. Chapter 3, *Romance, Suspense, and Translations*, describes the function of translations in the

Dutch literary field and explores new ways of measuring their linguistic characteristics. The chapter also goes into genre issues regarding the *Fifty Shades* trilogy and zooms in on the categories of Romance and Suspense. Chapter 4, *Literary Novels Written by Women*, first explores novels translated into Dutch and then those written in Dutch. The same approach can be found in Chapter 5, *Literary Novels Written by Men*. In Chapter 6, *Style, Gender, and Genre*, I present an overview of the results of the PhD-theses of Van Cranenburgh and of Koolen, addressing linguistic features and the role of gender. Additionally, I juxtapose the Dutch literary novels written by female and by male authors that were kept separate in chapters 4 and 5, to get a better view of their similarities and differences in writing style and perceived literary quality. In Chapter 7, *The Riddle of Literary Quality Solved?*, I come to a general conclusion, putting the results in a wider context and pointing out directions for follow-up research.

From Sociology of Art to Neo-Darwinism

The French sociologist and philosopher Pierre Bourdieu developed his very influential ‘field theory’ about how people value art, including literature. He deliberately did not pay attention to the exact shape and form of art, but focused on the sociological process of value attribution. Assigning value to literature and other artistic expressions, Bourdieu stated, is a highly complex task that amounts to deciphering a set of intricate and sometimes contradictory rules (1993, 217–218). The production and reception of literature take place in what Bourdieu called a ‘complicated struggle for power’. Not only writers play a role in the field but also those who acquire and publish literary works and those who influence public opinion, such as critics, researchers of literature, jury members of literary prizes, and teachers. The field is characterized by a precarious balance. A small change in one area will affect the entire field.

According to Bourdieu, the literary field as a whole has a clear hierarchy which can be seen as a continuum with two poles. At one pole, the subfield of large-scale production, the economic principle dominates. There, everything revolves around novels, stories, poems, etc. that are read, appreciated and, above all, consumed by a large audience. Authors and publishers who have chosen to operate in this part of the field are applying a strategy of minimal financial risk. The works produced and published appeal to the masses. They are sold in a short period of time and fill the coffers of authors and publishers with economic capital, but then usually quickly disappear from view. The other pole is the subfield of limited production. It publishes works of authors intended to be read and appreciated by like-minded people, the in-crowd, the peers: usually other writers and professional readers,

researchers, and teachers. A work appreciated by this group receives symbolic capital – not money, at least usually not enough in smaller language areas such as Dutch (although there are exceptions), but literary standing and prestige. Authors and publishers who position themselves in this part of the field do not opt for short-term security, but hope for long-term success: an ‘ever seller’ and perhaps inclusion in the literary canon. That may mean sufficient income for publishers, but usually not for the authors themselves. Bourdieu thus pointed out a number of fundamental oppositions. On the one hand there is the subfield of large-scale production, on the other hand that of limited production. Economic capital versus symbolic capital. Works that temporarily attract a lot of attention, bestsellers, versus timeless classics, ‘real’ literature. Authors of bestsellers earn lots of money. Authors of works of high literary quality are, by definition, struggling to make a living. But those who write for eternity do accumulate prestige and literary status, for example by being awarded literary prizes. The loser in economic terms is the winner in the race for literary prestige (1993, 39).

The economy of the subfield of limited production was characterized by Bourdieu as an anti-economy (1993, 54). He stated that economic success is even considered suspect. If a book sells sensationally well, people unconsciously deem it unlikely to be ‘high literature’, Bourdieu stressed. In the past, this assumption led to the exclusion of popular literature from scientific research, which only paid attention to the literary canon (1993, 39) – as Karel van het Reve noted in his famous talk in Leiden to which I referred earlier in this chapter. The subfield of large-scale production, that is, of bestsellers, is conservative, as Bourdieu called it. Authors and publishers who focus on this subfield are very much aware of what the public has already appreciated before. They try to deliver what fits the existing taste of the public. This does not mean that nothing ever changes, but innovations ultimately arise from developments at the opposite pole, the subfield of limited production. There, publishers and authors who strive for literary prestige actually try to break through existing traditions and consciously come up with new things. A younger generation tries to distinguish itself from the older generation in this way. The avant-garde does the risky experimenting and only those elements that have been proven to appeal to the wider public are taken over in the pole of large-scale production. This, in turn, causes the avant-garde to want to distance itself from this in a kind of push and pull movement (1993, 58). The consequence is that players at both poles – but for opposite reasons – constantly have to slightly adjust their strategy in order to obtain the success they are looking for, be it money or status, economic or symbolic capital.

It is clear why Bourdieu never commented on what this means in concrete terms for the texts themselves. His theory describes a process and not its tangible, in this case linguistic, output. In Bourdieu’s mind, the form of the texts entirely depends

on the context, and since the context is constantly changing, it is not surprising that the texts themselves are also constantly in flux. Moreover, as Bourdieu stated in the foreword to *Distinction* (1984/2010), the English translation of his book *La distinction* (1979) in reply to criticism, the precise workings of the sociological process can be slightly different in different countries and time periods. What mattered to him, however, was the underlying principle.

Moreover, Bourdieu stated that it is of main importance for authors to publish their work in the appropriate place to be considered of literary importance. They need to be visible in the right magazine and need to be published by the right publisher (1993, 95). Awards are important, but here too, the work must be awarded the correct prize. An award that would delight an author from the other pole may have a very negative impact on status and prestige on the other (1993, 100). The manuscripts that a publisher receives can be seen as the result of a preselection made by the submitting authors themselves. Authors have written their work keeping in mind which subfield they want to play a role in, and they approach those publishers who they assume to be active in that subfield. The same applies to critics. They receive books for review that have in turn been pre-selected by the publisher who published them. Review copies are sent to those who the publisher expects to pay (positive) attention to the titles, based on their previous reviews (1993, 133–135). Everything is connected to everything else. Opinions may seem very personal, but are in fact always collective judgements that are directly or indirectly related to positions in the literary field and to social class. Social conventions, in other words, are the basis for literary conventions. They are what perceptions of literary quality are embedded in.

Bourdieu maintained that the taste of readers depends primarily on the reader's level of education and after that, on their social class (the focus of Bourdieu 1984/2010). Bourdieu observed that especially readers who do not write themselves, but who do have a higher education and are members of the intellectual elite, allow themselves to be influenced by the unwritten rules of the literary field (1993, 41). He did not go into *how* readers develop their literary taste, however. This is where other sociologists picked up and developed what they called 'the new sociology of art'. They brought back into the picture the form and shape of the artworks themselves. Since this corresponds to how in the project *The Riddle of Literary Quality* we wanted to investigate what constitutes literary quality by looking at the texts themselves, their insights may be useful in this quest.

In 2007, sociologist Antoine Hennion theorized how 'amateurs', non-specialists, develop their taste and how they arrive at a value judgement. He illustrated his theory with the description of the ways in which different mountaineers try to reach the same summit. They have similar resources at their disposal, but not the same physical condition and climbing experience. Climbers therefore choose

different routes to reach the top. The mountain is not a social construction, Hennion emphasized, but it does have the potential to make differences between people visible. One can climb a mountain via all sorts of routes and still end up at the same summit, sharing the same view.

The same applies to taste. People can have the same taste and the same value judgement. But that doesn't mean that the judgement has come about in the same way. They may have developed their taste in different ways. Taste is not the result of unconscious and passive sociological processes, but is actively acquired by each individual. Taste is therefore a choice, Hennion stated. The preferences one has developed in the past determine who one is. What those preferences are depends on one's family, education, and the social environment in which one has functioned. In forming our opinions, almost all of us look at what others think and how they arrived at that judgement before we form our own. The sociological processes never completely disappear from view. But where Bourdieu looked at macro-processes in which taste was mainly seen as a characteristic of social class, Hennion concentrated on the analysis of processes at a micro-level, such as the ways in which art is advertised and how judgements about art are made.

Seeking to connect the macro (Bourdieu) and micro (Hennion) approaches, sociologist Ori Schwarz suggested in 2013 that new research should focus on 'tasting techniques' – ways in which taste is developed. Schwarz believed that mastering certain techniques in order to arrive at a value judgement can indeed be related to a person's social background and education. People from different social groups can thus have at their disposal different techniques for determining the value of a work of art – and ultimately arrive at the same judgement. Value judgements, in that sense, are not reducible to the influence of other people and their subjective experiences and preferences, but rather to the mastery of certain things, techniques for arriving at a judgement. At the same time, Schwarz argued, the 'right' way to consume art can be seen as a status symbol and an indication of one's identity – which agrees with Bourdieu's perspective.

Schwarz raised several issues that were ignored by Hennion. Firstly, some genres are valued significantly more than others, and this high level of appreciation is usually linked to an emphatic aversion to other genres. Often, we see a clear relationship between a genre and a particular age group or social class – there will be very little overlap between groups that appreciate classical music and heavy metal, Schwarz assumed. Furthermore, it is precisely the differences in the ways individuals develop their taste that can be related to social structures. Hennion's 'amateurs' clearly had characteristics of certain social classes. And thirdly, many people will have experienced how they themselves or others were excluded on the basis of their taste or the way in which they arrive at a value judgement about a work of art. Consciously or unconsciously, we judge someone's professional capacity

or social suitability on the basis of their way of life, taste, or mode of judgement. The nature of attention to cultural products is thus partly determined by social stratification, but conversely there is also an influence on the social structure. This influence is visible not only when it comes to identifying suitable candidates for a particular job, but also, for example, when it comes to finding friends or a life partner.

That last remark was prompted by psychological research with a (Neo-)Darwinian angle. Schwarz referred to a book published in 2010 by psychologist and linguist Paul Bloom, *How Pleasure Works: The New Science of Why We Like What We Like*. Bloom observed that people often attribute the reason for appreciation, pleasure, or whatever we want to call it, to properties of that which is enjoyed (sex, art, play) or consumed (food). Translating that to our research on literary quality, readers often seek the reason for the literary value they attribute to a text in the text itself. Bloom was convinced that this is only part of the story. He believed that enjoyment is influenced by other factors. It all starts with what we like or don't like to eat. This can be predicted if it is known what part of the world someone originates from. A predilection for something has to be learned. And sometimes people say they like something because it indicates, for example, that they have enough money to afford that food. In other words: taste as a status symbol.

Can we indeed compare reading literature to eating unpopular vegetables such as broccoli or Brussels sprouts? As an acquired taste? It takes a while to get used to the taste. And some people will never like them, but for the greater good – their own child's education, for example – they are willing to pretend every now and then that they absolutely love them. Bloom did not go into this, however. One possibility he did mention is that when people look at art (or, for that matter, read literature), they consider how difficult it is to make. The more effort involved in the creation, the greater the appreciation. But even that is not enough. More important, and according to Bloom even the essence, is that we think that the effort put into a work of art is only successful if it is done by a creative genius. Because that is what we have learned about artists. So, paintings by Vermeer are highly valued art, but those by Vermeer imitator Van Meegeren are not, even though he might have spent just as much time on them.

On this point, Bloom disagreed with Denis Dutton, who described the role art may have played in the evolution of mankind in the 2009 book *The Art Instinct: Beauty, Pleasure, & Human evolution*. According to Dutton, art is part of the process of sexual selection. He assumed that artistic men (yes, men...) are attractive partners for women and that, for this reason, creative genes are spread throughout the human population. It is difficult to create good art, Dutton argued, so an artist shows that he is clever, inventive, and can plan well. He is therefore a good partner. Bloom thought this short-sighted: it explains the attraction of the artist (the person) and not of the art (the object). He did not want to limit the sense of art to sexual selection from a

Darwinian perspective. After all, we are not constantly looking for a (genetically) suitable life partner; there are many other relationships that are important. We are also looking for friends, allies, and leaders. And so looking at what someone thinks about art can help in finding the right person for the right role.

To translate this to our subject: a shared literary taste can be a way of connecting with others. Take a look at my bookcase, and you will know how likely it is that we will get along or that there is work to be done. Literary taste as a social lubricant. But that still leaves the question of why one novel is 'allowed' on one's bookshelves and another is not. To what extent is symbolic capital – prestige – responsible for one's literary taste and to what extent is the attributed value inspired by how the text is shaped in language? In the project *The Riddle of Literary Quality* we harnessed new approaches developed in computational literary studies to answer this question.

Computational Literary Studies

The quantitative approach to text analysis firmly sits in the discipline of what in Europe has recently become known as computational literary studies. The field as a practice and as a community has been around much longer, with labels such as 'digital literary studies', 'stylometry', or 'distant reading'. The first use of 'computational literary studies' can be traced back to my professorial chair in computational literary studies at the University of Amsterdam, which was created in 2012. At the start the name was not used yet, nor in the first phase of the important COST Action Distant Reading for European Literary History (CA16204) which ran from November 2017 until May 2022 and was funded by the European Commission. It was introduced, however, in 2020 when the German priority programme SPP 2207 Computational Literary Studies was funded by the German Research Foundation (DFG). This priority programme aims to develop and establish innovative computational methods in the discipline of literary studies. The first phase of this programme funded eleven research projects. At the time of writing, the second phase of the programme was close to being launched. Another new project, developed as a follow-up to the COST Action Distant Reading, was named Computational Literary Studies Infrastructure. This project received funding from the European Commission in the Horizon 2020 framework INFRAIA-02-2020 (Integrating Activities for Starting Communities). It runs from March 2021 until March 2025. As stated on the website of the European Commission, the aim of this project is to 'help build the shared and sustainable infrastructure needed to undertake literary studies in the digital age (...).The resulting improvements will benefit researchers by bridging gaps between greater and lesser-resourced communities in computational literary studies and beyond, ultimately offering opportunities to create new research and insight

into our shared and varied European cultural heritage' ('Computational Literary Studies Infrastructure'). And in 2022 the first issue of the international, open access and peer-reviewed *Journal of Computational Literary Studies (JCLS)* was released, advocating itself on its website as responding to 'the increasing differentiation of subfields within the Digital Humanities, an ongoing process in which Computational Literary Studies has already gained considerable maturity and visibility.'

The project *The Riddle of Literary Quality* is rooted in new developments in stylometry starting with a measure called Burrows' Delta Procedure for authorship attribution and in the subsequent turn towards questions beyond authorship. Around the turn of the century, authorship attribution using computational means was a prominent topic in what was then called Humanities Computing. The assumption underlying this research is that comparison of a disputed text with a large corpus of texts written by authors who may be candidates for the authorship may help to identify the person who most likely was the real author of the disputed text. The corpus needs to consist of texts written in the same period of time as the disputed text and it must be certain who wrote them. Various methods can be used to examine the corpus. One of the most successful is the measure we now call Delta or Burrows' Delta, after the Australian literary scholar John Burrows (1928–2019), who developed this method and first published about it in 2002.

As I make extensive use of the Delta Procedure in this monograph, I will explain it in more detail than other methods that will be mentioned later in this chapter. In Burrows' original implementation, the Delta procedure builds a frequency list of the 150 most frequent words for each text in the corpus and for the selected corpus as a whole, listing the relative frequencies for each of the selected words. It then compares the frequency of each word in each text to the relative frequency of that word across the entire text collection. For all words, the deviations of the frequency from the mean are added up, resulting in a number for each text, the 'Delta'. That number expresses how much the total of all 150 most frequent words in a text differs in terms of frequencies – the number of times a word occurs in the text – from the occurrence of those words in the entire text collection. The author of the text with the Delta score closest to the score of the disputed work is most likely the author who should also be held responsible for that text.

Many experiments using texts of which the authorship was undisputed have shown that the Delta procedure works very well indeed. Using a much longer list of the most frequent words works even better, as David Hoover has shown in several contributions (e.g. Hoover 2004). There are some limitations, however. The size of the used texts does matter; the preferred minimum size of texts ranges between one thousand to five thousand words (tokens) (cf. Eder 2015). As in other authorship attribution methods, the Delta score only indicates the degree of difference in word use and word frequencies for the texts that are part of the actual corpus, which

means that the researcher's observations are only valid for that specific corpus. The procedure always yields a text that most closely resembles the disputed text, also if the real author is not represented in the corpus. Another issue to remember is that the method does not work if the disputed text is an author's very first publication, because in this case, there is simply no work available with which the disputed text can be compared. And finally, the procedure usually is only trustworthy if the corpus consists of texts belonging to the same genre. Experiments have shown that authorship attribution methods yield the best results if all texts in the corpus belong to the same genre, for example only literary novels, or only plays, and not a combination of both (cf. Burrows 279). This suggests that there are genre-dependent conventions in language usage that are responsible for differences in word frequency patterns.

With the launch of his Delta procedure, Burrows set in motion a whole new authorship attribution 'craze' in the digital humanities. Coming from a background in literary studies, he especially influenced literary scholars to use his method and further develop the procedure. The main reason why in computational literary studies Delta became such a huge success (next to its success rate in correctly identifying authors) is that it is transparent: it deals with words, interpretable linguistic units, and these words and their frequencies are relatively easy to inspect and as such help the researcher to get a closer look at the data, to explore the texts in more detail and to uncover the directions in which to continue the research on the way to a satisfying and fruitful interpretation. Delta helps us to creatively think about our textual data and how they function. Delta was implemented in the Stylo Package for R developed by Maciej Eder, Jan Rybicki, and Mike Kestemont and in this form reached a whole new audience of scholars who benefitted from the graphical user interface that made using Stylo and applying Delta a lot easier. Eder and the other members of the Computational Stylistics Group have been teaching how to use this package all over the world for the last decade or more. Their website lists the ongoing work and the many publications related to Stylo. I have benefitted greatly from their work, and a lot of my own research as presented in this book is done with the Stylo package; I will go into more detail about the different measures I applied in the next chapters of this book.

What Delta has taught us is that by using digital text corpora and statistical measures we actually can recognize patterns in language use that are usually invisible to a human reader. Surprisingly, authors may be recognized by their word frequencies. And the biggest surprise is that it is mainly the very common words that betray the identity of an author. Words that appear in every text and in the works of every author. Articles, conjunctions, prepositions: no text goes without them. If we instruct software to count all the words in a text and compare the results with the number of occurrences of the same words in other texts, it turns

out that there are subtle differences in how often different authors use the same very frequent words. For example, one author chooses an article just a little more often than others. Another author uses more adjectives compared to other writers. Those differences may, of course, indirectly depend upon the topics authors deal with; they could also result from deliberately chosen and divergent approaches. We cannot yet determine exactly how it works, because there are so many different variables involved. For the time being, however, most researchers assume that the differences can partly be traced back to the unique development of the language skills of each individual author, as David Hoover, one of the scholars who has done a lot of work enhancing the Delta procedure, writes in his most recent book *Modes of Composition and the Durability of Style in Literature* (3–8).

A logical next step after identification of individual authors was to apply authorship attribution models such as Delta to find out more about collaboration between authors. I will mention some examples from different languages, genres, and time periods. Joris van Zundert and I analysed a Medieval Dutch Arthurian novel that was supposedly left unfinished by the first author and was completed by another. We used a walking window version of Delta to find out where the second author took over from the first. Mike Kestemont, Sara Moens, and Jeroen Deploige explored how Hildegard von Bingen and her different amanuenses, all writing in Latin, collaborated. I looked into the collaboration of two eighteenth-century Dutch female authors who collaborated on four epistolary novels to see if critics denying one of the two any literary talent were correct or if we should give precedence to what the two women themselves stated again and again, that they collaborated on an equal footing. I also considered if the two authors created a character voice for each of their letter writing characters or not. Lisa Pearl, Kristine Lu, and Anousheh Haghighi followed up on this research with their computational analysis of character voice in Samuel Richardson's epistolary novel *Clarissa*. Another interesting study on assumed collaboration was published by Rachel McCarthy and James O'Sullivan, testing an assumption of some readers that Emily Brontë's *Wuthering Heights* must have been co-authored by her brother Branwell. For the modern time period, Jan Rybicki, David Hoover, and Mike Kestemont, for example, focused on works co-authored by Joseph Conrad and Ford Madox Ford trying to find out if Ford was indeed the author of a sizeable fragment of *Nostromo*, as he claimed. Jan Rybicki and Magda Heydel analysed the nature of a translation into Polish of a novel by Virginia Woolf that was started by one translator and finished by another.

The new methods for authorship attribution as sketched above also inspired scholars to ask questions that go beyond authorship. An example is the analysis of the chronological development of one author's style using the same measures as for authorship attribution (Hoover 2021, 122–142). One of the first researchers of

literature to look beyond verification of authorship was Matthew L. Jockers. In 2013 he published *Macroanalysis*, a monograph in which he reported on his computational research into a corpus of around 3,500 English-language novels, all first published in the nineteenth century. Jockers was interested in measuring such things as genre, origin, and influence. Can certain differences in style be traced back to, for example, the time period in which a work was written, or to the genre to which it can be said to belong? Is the geographical origin of the author – in addition to American works Jockers also included novels by English and Irish authors in his research – recognizable in the writing style? Do other external factors play a role in the design of a novel? And can we perhaps even find out which authors most influenced others, looking only at the digitized texts themselves?

Jockers found that the overall style of the novels did not change very much throughout the nineteenth century. What did change over time was the presence of certain kinds of novels, genres that are recognizable by, among other things, shared topics and an associated shared style. Jockers concluded that genres develop and disappear over a period of about thirty years, which agrees with what other researchers have found (84). He also established that some authors have a much more varied style in their different works than others. He noted that the style of Jane Austen, for example, was very easy to recognize with computational means. All her novels are similar in terms of the use of the most frequent words; the differences only show in the topics of the narratives, which are reflected by less frequent content words. In contrast, Mark Twain and James Joyce offered such a wide variety of styles that the software had the greatest difficulty (so to speak) in recognizing these authors (93). Jockers also used the software he developed to predict the author's gender based on the texts themselves. This proved to be correct in eighty per cent of the cases. He was particularly interested in the 20 per cent where things go wrong (93–95). All in all, Jockers' methods worked best in recognizing differences in authorship. Then, in descending order of correct classification, followed genre, decade of publication, and finally gender of the author (97).

Jockers also tried to delve deeper in terms of content. Can software help to discover themes in the novels? Which topics are characteristic of certain genres or authors or countries? For this he used a technique called topic modelling, which determines for the selected text collection which words often occur in close proximity (more about topic modelling in Blei 2012). The software produces a number of word lists, to which the researcher then has to apply a label that describes the subject connecting the words. An example from Jockers' topic modelling is the list 'heart, cabin, bog, county, mountains, Derry, house, air, evening, turf, friends, heaven'. Jockers labelled this list of co-occurring words as the topic 'Ireland' (124). With these lists of words, the proportion of the topics was measured in all novels. The results showed how they are distributed across

different authors, genres, time periods, origin, or gender of the author. In his nineteenth-century English texts Jockers found, for example, that the theme 'Villains and Traitors' occurred more often in novels by male than by female authors, just like 'Quarrels and Duelling' and 'Enemies'. Female authors paid more attention to 'Affection and Happiness' and 'Female Fashion'. The topic of 'American Slavery' appeared most often in novels from the United States and less so in those from Great Britain and Ireland, although the slavery discussion there was equally intense. This became apparent when Jockers compared the results with those of the theme 'Native Americans', where the imbalance was much greater between novels from the United States on the one hand and those from Ireland and Great Britain on the other (136–141).

In 2016, together with Jodie Archer, Jockers published the book *The Bestseller Code*. Archer and Jockers now focused on modern novels, written in English or translated into English. More than fifty thousand novels are published in the United States every year. Around two hundred titles from that annual crop end up on the prestigious *New York Times* bestseller list, less than half a per cent of all published fiction. Archer and Jockers wanted to find out what the books that made it to the bestseller list have in common, hoping this would help them learn more about what makes a novel so good. For their analysis of the novels themselves, they applied the same collection of computer programmes that Jockers developed for *Macroanalysis*. The results helped to uncover which characteristics a novel should have to be bought and devoured by many readers, a 'bestseller code'. At least 30 per cent of a bestseller consisted of one or two main topics. The more different topics, the less well the novel sold. The narrative did not need to have a lot of sex, drugs, and rock and roll. Yes, sex sold, but only in a niche market, with a relatively small group of adepts. Part of the bestseller code was that a novel must be about relationships between people – friendship, enmity, love, hate, and everything attached to that. And those relationships must necessarily even be one of the main topics of the novel. The authors of Archer's and Jockers' bestselling corpus who best addressed this in their novels were Danielle Steel and John Grisham. These two also applied another rule perfectly: one of the topics addressed in a novel must be dominant and sort of serve as the glue for the story that is being told. Steel is known for her novels in which women find their way to fulfilled lives with meaningful relationships. Grisham's books revolve around characters with legal problems and how those problems turn their lives upside down. Realism was important in bestsellers, so topics like work, school, media, doctors, death, marriage, and taxes did well (52–68).

Archer and Jockers calculated the varying proportion of words expressing positive or negative emotions for each novel and graphed their use throughout a novel. They suggested this may present an impression of the plot, the storyline of a book (97). They looked for novels in their corpus that had a similar line to E.L. James's

hugely successful *Fifty Shades of Grey* (the Dutch translation of this novel will be mentioned quite often in the next chapters of this book). Bestselling novels with a comparable emotion plot are *Inferno* by Dan Brown, *The Circle* by Dave Eggers, and *The Cuckoo's Calling* by Robert Galbraith aka J.K. Rowling (110–111).

The style of the novels also was very important, stated Archer and Jockers (more about the definition of style in a moment). After all, the writing style is the building material with which authors shape the storyline, theme, and characters of their novels. The style of Jane Austen, which Jockers already explored in *Macroanalysis*, found a modern equivalent, according to Archer and Jockers, in the box office success novels of Jacky Collins (121–124). High-frequency words or parts of speech could be an important signal of bestseller potential, including the article 'the', Archer and Jockers noted. Furthermore, punctuation played a role. Proper use of periods and commas was important. Using question marks was fine, but an ambitious author should avoid scattering exclamation marks throughout a novel (131–136).

Archer and Jockers' *New York Times* best-selling collection included significantly more novels written by male authors than by female authors. However, when the researchers sorted the list by best-selling style, then the top of the list mostly consisted of female authors. Not just established authors like Steel, but also first novel authors. They did not know how to explain this. Interestingly, the computer analysis of modern novels proved more difficult to correctly predict the author's gender than it was for the nineteenth-century novels that Jockers examined in *Macroanalysis*. The success rate was now 71 instead of 80 per cent (137–138). When Archer and Jockers looked into the errors more closely, they found something very interesting. Well-known female authors such as Toni Morrison and Barbara Kingsolver were mistaken for men by the software. What these female authors and the male authors of a similar style appeared to share was an academic background, being university educated, or working as a scholar in academia. The authors of the books that were attributed to women by the software – sometimes incorrectly, including the first part of Stieg Larsson's *Millennium* trilogy – shared a background in journalism. In short, the interesting finding was that 'masculine' and 'feminine' style appear to be related to differences in education and occupational environment rather than differences in biological sex. Of course, it remains worrisome that this high level of education is apparently more often found in published and best-selling men than women (140–144).

Archer and Jockers analysed the titles of all the bestsellers in their corpus and saw several similarities. The word *girl* was very present, not only in the titles, but also in the novels themselves, which had a lot of young female characters (147–149). In the texts themselves, verbs expressing action were clearly an advantage. The bestsellers contained many common, everyday words and things. Dialogue was indicated with as few words as possible, so *ask* or *said* but not *exclaimed* or *demanded* (160–168).

In the closing chapter of their book, Archer and Jockers confessed that their research has changed their view on reading, writing, and literature. Archer, a researcher but also working in the publishing industry, was sceptical about the computational methods Jockers wanted to apply to the literary canon. She became convinced of their value. Jockers was surprised that Archer wanted to explore popular contemporary fiction. He later added this component to his literary studies classes, convinced that his students should be exposed not only to 'literature' but also to modern popular fiction (200–201). I find it very interesting that both authors in fact admitted that their view as a reader was limited and that it had been pleasantly broadened by the outcome of their own research. This is a very relevant observation for our research into the riddle of literary quality: individual reading experiences are of great importance for what new generations of students are taught, to name but one aspect. It seems that unconscious processes can have a major influence on how and what readers read and how the value of the novels is being determined.

In our project *The Riddle of Literary Quality* we clearly went beyond authorship in asking ourselves the question: Can we identify differences in word frequencies (and much more) between books that are perceived as highly literary by their readers and those novels that are not considered to have any literary quality? And whatever the answer, what does that teach us about what we call literature? Is it just a question of style? Readers unfamiliar with authorship attribution and computational literary studies often scratched their heads when we referred to style when we spoke about the relative frequencies of articles and prepositions and other most frequent words in our analyses of digital text corpora. This is understandable. If we look at what has been understood by the term 'style' in Dutch, French, and German literary studies over the past seventy years, we see differing views about what style is and how it can be researched – or not. For example, there is the view that style embodies an idea of beauty that can only be judged emotionally. Another view is that style is the result of a choice an author makes from the various possibilities offered by a language. A third example is style as an exception to the norm: only 'special' language use is relevant when considering writing style, such as some, but not all, similes or metaphors. The idea that style is also ingrained in the way 'meaningless' words are used by an author has only come into play since we can use software and digital texts to count much more and much faster than ever before (Herrmann et al. 30).

I want to illustrate this with the results of research done by Erik Ketzan and Christof Schöch on Andy Weir's popular novel *The Martian* from 2014, launched as a successful film in 2015 (Ketzan & Schöch 2017, Ketzan & Schöch 2021). Set in the year 2035, the story revolves around Mark Watney, an American astronaut and botanist who, by an unfortunate accident, must survive alone on the planet Mars.

In his log we read how he manages to do that. *The Martian* is the first published novel written by Weir. Because he did not find a publisher, he first made the novel available in instalments on his own website in 2011. The story became so popular that fans asked Weir to publish it as an Amazon Kindle book. When *The Martian* then was downloaded even much more often, publishers finally became interested and this led to the book being taken up by an established publishing house.

Before the book publication, the story was carefully checked and subjected to a meticulous editing process. Ketzan and Schöch applied software to investigate the differences between the two versions of Weir's novel. In the book edition, the story had not changed, the characters had remained the same, and the structure of the text had also been untouched, except for the omission of the original epilogue. The changes they found were minor in themselves, but they totalled more than five thousand. Spelling errors were corrected, as were misuse of hyphens or capitalization. Abbreviations were written in full. Most numbers were converted from Arabic numerals to words (8 became eight, for example). Another common change in the book edition is the removal of a lot of expletives or the replacement of such statements by a somewhat milder version. The most significant change is the omission of an epilogue in which Watney, back on Earth, behaves horribly towards a fan. The concluding chapter of the book edition contains an addition in which Watney expresses gratitude to all who contributed to his rescue and emphasizes his faith in humanity.

Ketzan and Schöch (2017) argued convincingly that all these small changes – and the one somewhat bigger one – did indeed substantially change the ‘tone’ of the novel. They have led to changes in theme and style. The many numbers and spelling errors in the log of protagonist Watney, for example, fit within a realistic depiction of his extreme experiences. Straightening those out makes the log less realistic, turning the ill-fated protagonist who takes notes only for himself into someone who is mindful of eyes accustomed to reading perfectly edited texts. (And for those who have the urge to say that Watney records his spoken log in the film and therefore cannot make any spelling mistakes: one of the first sentences of the book is ‘I don’t even know who’ll read this’, which absolutely suggests written or typed text.) It seems very likely, Ketzan and Schöch (2017) argued, that the purpose of all those minor changes was to make the main character more relatable and sympathetic and thus make the novel *The Martian* more interesting for an even larger audience.

The thousands of small changes of seemingly quite meaningless language aspects have indeed changed the style of *The Martian* substantially. And that is why we needed a new definition of style. Together with my colleagues from Germany Berenike Herrmann and Christof Schöch, I have proposed a new definition of ‘writing style’: Style is a property of texts constituted by an ensemble of formal

features which can be observed quantitatively or qualitatively (Herrmann et al. 44). In particular the word 'quantitatively' has acclimatized the concept of style to the computational methods currently available for literary research.

Empirical Study of Literature

Computational literary studies was not the only research discipline that inspired our research. A volume edited by Willie van Peer was one of the other inspirations that led to the project proposal for *The Riddle of Literary Quality*. Van Peer was one of the founders of the discipline of the empirical study of literature. Researchers active in this area mainly use methods from psychology, designing survey-based experiments, targeting respondents, and presenting statistical analysis of the results. For example, researchers want to investigate how people read, what expectations they have of the texts they are confronted with and how those expectations may influence their way of reading. Some other recurring research questions are: what is the relationship between reading and empathy? How does literature play a role in the mental health of readers?

The research partly harks back to formalism and structuralism, literary theories concentrating on the textual side of literature developed in the first half of the twentieth century. In formalism, the leading idea was that the literary nature of a text had to do with striking, deviant use of language. Structuralism took a more nuanced approach and explored the idea of deviation more in context. Theorists spoke of 'foregrounding' as a distinctive feature of language use in a literary work, by which they meant that a text stands out due to a deviating, creative use of stylistic devices and thus brings itself to the attention of the reader (Bertens 24–40). In empirical studies of literature the idea of 'foregrounding' is still strongly present. Empirical scholars became convinced, however, that a text needs more than striking and deviant use of language to be perceived as having literary quality. Some experiments, for example, showed that sentences with unexpected use of language take a reader slightly longer to read than sentences without such surprises (see for example the article 'How Literary Can Literariness Be? Methodological Problems in the Study of Foregrounding' by Massimo Salgaro). But what surprises one reader may be normal to another, or may come to feel normal to the same reader over time – and may stop being surprising. What is unexpected, therefore, strongly depends on the context and on the reader. In short, the individual experiences of readers play an important role, as do their expectations (Stukker and Verhagen 56).

The inspirational collection of studies van Peer edited, *The Quality of Literature: Linguistic Studies in Literary Evaluation*, was published in 2008. The volume addressed a variety of ways to investigate the language use of texts in search of

knowledge about what may qualify a text as having literary quality. In his introduction to the collection, van Peer emphasized how unusual this approach was. Until then, literary scholars, following Bourdieu, mainly studied the cultural context and also the ways in which readers react to literary works, but hardly the text itself. He hoped the volume would help to change that. In his own contribution to the volume, 'Canon Formation: Ideology or Aesthetic Quality?', he invited scholars to pay more attention to texts that are no longer read. The method van Peer advocated in this contribution is contrasting well-known literary works with forgotten ones and exploring their textual properties. He intuited that many different factors play a role in what makes a text literary. He was convinced that literary quality is also a property of the text itself and cannot be explained solely by the sociocultural context or by differences between readers.

Conclusion

In this introductory chapter I addressed the main question of *The Riddle of Literary Quality*: can we measure the literary value of novels in the texts themselves when we correlate the computational analysis of linguistic features with readers' perceptions of literariness of the same texts? This innovative mixed-methods approach on the one hand took its cue from the empirical study of literature, with its focus on actual readers, and on the other hand built on recent developments in computational literary studies in its shifting emphasis from authorship attribution to issues going beyond authorship. I presented some of the most relevant publications from this discipline in a bit more detail, not mentioning many other highly relevant and interesting articles, chapters, and monographs that could surely provide even more context and insights into the affordances of computational approaches for literary studies. Some of the most recent books providing important contributions to the field were published by Sarah Allison, Katherine Bode, José Calvo Tello, Martin Eve, Andrew Piper, and Ted Underwood and can be found in the bibliography of this chapter; their monographs and the many publications they refer to in their work will undoubtedly help interested readers get even better acquainted with the field. I also zoomed in on the Netherlands as the geographical area of which the conventions of literary quality are my topic of analysis. I sketched the Netherlands' place in the periphery of the 'world republic of letters' and summarized the field theory developed by Bourdieu and the new sociology of art that it gave rise to. Bourdieu's work has been highly influential in Dutch literary scholarship even until today and in fact is the theory that is being challenged by the approach we selected for *The Riddle of Literary Quality*. The sociological part of the project mainly revolves around the large survey we developed and ran in 2013 and which is the topic of the next chapter.

References

All hyperlinks were last accessed in February 2023.

- Allison, Sarah. *Reductive Reading: A Syntax of Victorian Moralising*. Johns Hopkins University Press, 2018.
- Archer, Jodie, and Matthew L. Jockers. *The Bestseller Code*. Penguin Books, 2016.
- Bertens, Hans. *Literary Theory: The Basics*. 2nd ed., Taylor & Francis, 2008.
- Beyers, Ton, et al. *Nederlandse kunst in de wereld: Literatuur, architectuur en beeldende kunst 1980-2013* [*Dutch Art in the World: Literature, Architecture and Visual Arts 1980-2013*, not available in English]. Vantilt, 2015.
- Blei, David M. 'Probabilistic Topic Models'. *Communications of the ACM* vol. 55, no. 4, 2012, pp. 77–84, doi:10.1145/2133806.2133826.
- Bloom, Paul. *How Pleasure Works: The New Science of Why We Like What We Like*. Norton, 2010.
- Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary History*. University of Michigan Press, 2018, doi:10.3998/mpub.8784777.
- Bourdieu, Pierre. *Distinction: A Social Critique of the Judgement of Taste*. 1984. Translated by Richard Nice. Routledge, 2010.
- Bourdieu, Pierre. *The Field of Cultural Production: Essays on Art and Literature*, edited by Randal Johnson. Polity, 1993.
- Brems, Elke, et al., editors. *Doing Double Dutch: The International Circulation of Literature from the Low Countries*. Leuven University Press, 2017.
- Burrows, John. "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing*, no. 17, 2002, pp. 267–287, doi:10.1093/lc/17.3.267.
- Calvo Tello, José. *The Novel in the Spanish Silver Age. A Digital Analysis of Genre Using Machine Learning*. Bielefeld University Press, 2021, doi:10.14361/9783839459256.
- 'Computational Literary Studies Infrastructure.' *European Commission Community Research and Development Information Service (CORDIS)*, 28 April 2022, <https://cordis.europa.eu/project/id/101004984>.
- Computational Stylistics Group, <https://computationalstylistics.github.io/>.
- 'Distant Reading for European Literary History.' European Cooperation in Science and Technology (COST Action CA16204), <https://www.cost.eu/actions/CA16204/>.
- Dutton, Denis. *The Art Instinct: Beauty, Pleasure, and Human Evolution*. Oxford University Press, 2009.
- Eder, Maciej. 'Does Size Matter? Authorship Attribution, Small Samples, Big Problem'. *Digital Scholarship in the Humanities*, vol. 30, no. 2, 2015, pp. 167–182, doi:10.1093/lc/fqt066.
- Eder, Maciej, et al. 'Stylometry with R: A Package for Computational Text Analysis'. *The R Journal*, vol. 8, no. 1, 2016, pp. 107–121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.

- Eve, Martin Paul. *Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell's Cloud Atlas*. Stanford University Press, 2019.
- Franssen, Thomas. *How Books Travel: Translation Flows and Practices of Dutch Acquiring Editors and New York Literary Scouts, 1980-2009*. 2014. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245/1.446961>.
- Hennion, Antoine. 'Those Things that Hold Us Together: Taste and Sociology.' *Cultural Sociology*, vol. 1, no. 1, 2007, pp. 97–114, doi:10.1177/1749975507073923.
- Herrmann, Berenike J., et al. 'Revisiting Style, a Key Concept in Literary Studies'. *Journal of Literary Theory*, no. 9, 2015, pp. 25–52.
- Hoover, David L. 'Testing Burrows's Delta'. *Literary and Linguistic Computing*, vol. 19, no. 4, 2004, pp. 453–475, doi:10.1093/lc/19.4.453.
- Hoover, David L. *Modes of Composition and the Durability of Style in Literature*. Routledge, 2021.
- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013.
- Journal of Computational Literary Studies* (JCLS), <https://jcls.io/>.
- Kestemont, Mike, et al. 'Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux'. *Digital Scholarship in the Humanities*, vol. 30, no. 2, 2015, pp. 199–224. doi:10.1093/lc/fqt063.
- Ketzan, Erik and Christof Schöch. 'What Changed When Andy Weir's *The Martian* Got Edited?' *Digital Humanities Conference 2017*, Montréal, 8–11 Aug 2017. Conference Abstract, pp. 285–288.
- Ketzan, Erik and Christof Schöch. 'Classifying and Contextualizing Edits in Variants with Coletto: Three Versions of Andy Weir's *The Martian*'. *Digital Humanities Quarterly* (DHQ) vol. 15, no. 4, 2021, <http://digitalhumanities.org:8081/dhq/vol/15/4/000579/000579.html>.
- Koolen, Corina. *Reading Beyond the Female: The Relationship Between Perception of Author Gender and Literary Quality*. 2018. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37fice7>.
- McCarthy, Rachel and James O'Sullivan. 'Who wrote *Wuthering Heights*?' *Digital Scholarship in the Humanities*, vol. 36, no. 2, 2021, pp. 383–391, doi:10.1093/lc/fqaa031.
- 'Novel Perceptions: Towards an Inclusive Canon Project Website.' *University of Wolverhampton*, <http://novel-perceptions.thememorynetwork.com/>.
- Pearl, Lisa et al. 'The Character in the Letter: Epistolary Attribution in Samuel Richardson's *Clarissa*'. *Digital Scholarship in the Humanities*, vol. 32, no. 2, 2017, pp. 355–376, doi:10.1093/lc/fqw007.
- Piper, Andrew. *Enumerations: Data and Literary Study*. University of Chicago Press, 2018.
- 'The Riddle of Literary Quality Project Website.' *Huygens Institute, Royal Netherlands Academy of Arts and Sciences*, https://literaryquality.huygens.knaw.nl/?page_id=179.

- Rybicki, Jan, and Magda Heydel. 'The Stylistics and Stylometry of Collaborative Translation: Woolf's *Night and Day* in Polish'. *Literary and Linguistic Computing*, vol. 28, no. 4, 2013, pp. 708–717, doi:10.1093/lc/fqt027.
- Rybicki, Jan, et al. 'Collaborative Authorship: Conrad, Ford and Rolling Delta'. *Literary and Linguistic Computing*, vol. 29, no. 3, 2014, pp. 422–431, doi:10.1093/lc/fqu016.
- Salgareo, Massimo. 'How Literary Can Literariness Be? Methodological Problems in the Study of Foregrounding.' *Scientific Study of Literature*, vol. 5, no. 2, 2015, pp. 229–249, doi:10.1075/ssol.5.2.06sal.
- Schwarz, Ori. 'Bending Forward, One Step Backward: On the Sociology of Tasting Techniques'. *Cultural Sociology*, vol. 7, no. 4, 2013, pp. 415–430, doi:10.1177/1749975513480957.
- Stukker, Ninke, and Arie Verhagen. *Stijl, taal en tekst: Stilistiek op taalkundige basis* [Style, Language and Text: Linguistics-Based Stylistics, not available in English]. Leiden University Press, 2019.
- Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019.
- Van Cranenburgh, Andreas. *Rich Statistical Parsing and Literary Language*. 2016. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37fice7>.
- Van Dalen-Oskam, Karina, and Joris van Zundert. 'Delta for Middle Dutch: Author and Copyist Distinction in *Walewein*'. *Literary and Linguistic Computing*, vol. 22, no. 3, 2007, pp. 345–362, doi:10.1093/lc/fqm012.
- Van Dalen-Oskam, Karina. 'Epistolary Voices: The Case of Elisabeth Wolff and Agatha Deken'. *LLC: The Journal of Digital Scholarship in the Humanities*, no. 29, vol. 3, 2014, pp. 443–451, doi:10.1093/lc/fqu023.
- Van Dalen-Oskam, Karina. *Het raadsel literatuur: Is literaire kwaliteit meetbaar? [The Riddle of Literature: Can We Measure Literary Quality?]* Amsterdam University Press, 2021, <https://karinavdo.github.io/RaadselLiteratuur/>.
- Van Peer, Willie, editor. *The Quality of Literature: Linguistic Studies in Literary Evaluation*. John Benjamins Publishing Company, 2008.
- Van het Reve, Karel. 'Literatuurwetenschap: het raadsel der onleesbaarheid' [Literary Theory: The Riddle of Illegibility, not available in English]. Johan Huizinga Lecture, 8 December 1978, Leiden, https://www.dbnl.org/tekst/reve003lite02_01/index.php.

2. The National Reader Survey

Books, Readers, Opinions

Abstract: This chapter describes how the 2013 Dutch National Reader Survey was organized and what the main outcomes were, in such a way that the survey can be adapted for use in other countries or language areas. It presents the criteria for the selection of the four hundred plus one novels that participants could rate and zooms in on the four main categories of books on the list. This is followed by a description of the kinds of readers that took part and which books received the highest and lowest mean scores for general quality (how good the book was) and literary quality. It elaborately quotes from the motivations readers gave for their ratings and analyses the main trends by applying keyword analysis.

Keywords: Literary quality, Survey Design, Reader Roles, Keyword Analysis, Computational Literary Studies, Distant Reading

The Survey

The novelty of *The Riddle of Literary Quality* is the combination of computational text analysis methods with the analysis of readers' opinions about the literary quality of the same texts. The aim was to find out whether readers of certain novels agree or disagree about the literary quality of these novels, how large the variety of opinions may be, and whether particular trends or patterns may be observed. Another goal was to explore whether different kinds of readers or reader roles may be distinguished. Last but not least, the Riddle team was interested in how readers would motivate their opinions about literary quality when asked to describe the reasons for their ratings. The necessary data were gathered in 2013 in a large online survey – a logical choice, because in that year 95 per cent of the Dutch population already had access to the Internet.¹

¹ As stated by the Dutch Central Bureau for Statistics (CBS), <https://www.cbs.nl/nl-nl/achtergrond/2013/48/eenpersoonshuishoudens-maken-inhaalslag-met-internet>.

This chapter starts with a description of the composition of the survey and continues with a discussion of the selection of the corpus. It then presents a general overview of the results in more detail than was done in the 2020 article ‘Literary Quality in the Eye of the Dutch Reader: The National Reader Survey’ written by Corina Koolen, Karina van Dalen-Oskam, Andreas van Cranenburgh, and Erica Nagelhout. The outcomes that will be discussed inform the organization of the rest of this monograph, as will be made clear throughout the chapter and summarized at the end. For the exact numbers, percentages, and more details concerning the outcomes to be discussed, scholars may refer to the ‘litRiddle’ package for R that includes all data that were gathered in the survey (see Appendix 3).

Specialized in economic psychology, market researcher Erica Nagelhout helped the Riddle team to design and carry out The National Reader Survey (Het Nationale Lezersonderzoek), as we named it. The agency that took care of the technical implementation to present the survey to a wide audience also provided the option to hire participant panels representative of the Dutch population in terms of gender, age distribution, educational level, and place of residence. We used such a representative panel twice; the first time to test the survey before we finalized it, and the second time just before we closed the survey to the public, to be able to compare the information collected so far with the results from the representative panel of around five hundred inhabitants of the Netherlands. This would help us to see how the participants in The National Reader Survey compared to the Dutch population as a whole. The final version of the survey can be found in Appendix 1.

We asked respondents for their age, how many books they estimated to read per year, and whether they read only fiction, only non-fiction, or both. We also asked for their gender, highest level of education, and the first part of their postal code.² To learn more about possibly different types of readers having different opinions about the same book, the survey included sixteen statements about reading and participants were asked to what extent they agreed with each. In addition, they were asked to indicate to what extent they considered themselves to be *literary* readers.

Participants were presented with a list of the author’s name and title of 401 books and were asked to tick the check boxes of all books they could remember reading. Next, the survey software randomly selected up to seven of the books marked as read and then participants were asked to rate each of those books on two scales: in terms of general quality (how good or bad they thought the novel was) and in terms of literary quality. Score 1 was the lowest score and stood for

² Dutch postal codes consist of four digits representing an area, followed by two letters indicating a street or part of a street in that area. For privacy reasons, we were only allowed to ask for the four digits and not for the additional two letters.

'very bad' or 'definitely not literary' and 7 was the highest, 'very good' or 'highly literary'. Each also had an eighth option 'Do not know'. Participants could choose to rate seven more books and repeat this as many times as they wished for the books they had read. The survey software then selected one book that the participant had rated relatively high or low for literary quality compared to the other books that had been rated by the same person. For that book the respondent was asked: 'Can you explain why you rate this book this way?' A free field gave them ample room to motivate their score. Next, participants were shown the list of books without the ones they had previously indicated to have read. They were asked whether there were any books on the list that they had an opinion about, even without having read them. From the ticked boxes on this list, the survey software randomly selected up to seven books to be rated in the same way as the books they had read.

Some of the first participants were rather surprised by the question about books they had not read. This led the Riddle team to post a list of frequently asked questions on the National Reader Survey website, including 'Why do you also ask for opinions on unread books?' Our answer was: 'Sometimes people deliberately choose not to read a book. This suggests they do have an opinion about it. We would like to know that opinion.' This was one of the two questions we received most often. The other question was 'What do you mean by literary?' Our answer was: 'We would like to find out what readers consider literary. Many readers do have an idea about that. That's why we don't give a description of what we ourselves consider "literary".'

The National Reader Survey was launched on 4 March 2013 and the press release announcing it was widely picked up. Mentions appeared on the websites of various public libraries, publishing houses, and other book- and reading-related sites with a link to the survey. Some newspapers published an interview. When we noticed that proportionally fewer participants came from the Dutch provinces of Friesland and Limburg than from the rest of the Netherlands, we decided to launch two short promotional videos in Frisian and Limburgian, in order to attract even more respondents from those areas. With a message on the web magazine *Mixed Grill* we tried to get more men to participate. And to get more opinions about romantic-humorous books we advertised on the Dutch site *chicklit.nl*, at the time of writing in July 2022 still a very lively Dutch website, although the label 'chick lit' has generally fallen out of fashion because of the inherent sexism of the name.³

3 Kyra Schiffers wrote a communication plan and went in search of opportunities to generate attention for the National Reader Survey. Bas Doppen designed a front page for the website that led to the survey, featuring the logo of an open book with the words 'What do you think of me' above it. Renske Siemens ensured, among other things, that a bookmark with the logo was made. 250 copies of these were delivered by NBD Biblion, an organization supporting Dutch libraries, to every public library in the Netherlands.

After two weeks, we had received 635 completed surveys. After a month, the score was 1,078. Then the CPNB, the Foundation 'Collective Promotion for the Dutch Book', which aims to encourage people to read more books, sent out a mailing in which The National Reader Survey was prominently mentioned and a week later (16 April 2013) the counter suddenly stood at 7,480. From that moment on, the number of completed surveys went up steadily, with additional increases following a mailing from two bookstore chains and an advertisement at the check-out counter of the website *Boekwinkeltjes.nl*, where private book owners offer books for sale. On 20 June we had received 9,227 completed surveys and on 8 July 11,283. On 27 September 2013, when 18,484 people had started the questionnaire, we closed the survey. The number of participants who stopped prematurely was higher than expected, 4,700. In the end, 13,784 participants had completed the survey.

The Books

We were interested in what publishers call 'longer fiction' and decided that the survey corpus should include only novels or separately published stories of a certain length (with a minimum of twenty pages). To make relevant observations, we wanted to collect at least fifty ratings for each title. We calculated that we needed four hundred books.⁴ To ensure that our personal bias about which novels have or do not have any literary value would not inform our selection of the books we drew up an additional list of strict selection criteria. One of these related to the first time of publication. Because we were interested in the personal opinions of participants, unaffected (if possible) by what they had learned during their education, we selected only recently published fiction. We chose the year 2007 as the cut-off date, excluding books published before 2007. As explained in Chapter 1,

Johan Kwantes made two short films with authors Renate Dorrestein and Christiaan Weijts in which they asked readers to participate in the survey. The films to specifically recruit Frisian and Limburg participants were recorded by Ilja Nieuwland (Frisian) and Steven Surdèl (Limburgian).

4 We wanted to collect opinions on two scales: literary quality and general quality. Our initial assumption was that this should get opinions about four groups of texts: (1) books perceived as both literary and good; (2) books perceived as literary but bad; (3) books not perceived as literary but perceived as good; (4) books not perceived as literary and perceived as bad. In addition, we wanted opinions not only about books that had been read by the respondents, but also about books that had not been read. This resulted in a total of eight categories: read books that were considered literary and good, books not read that were perceived as both literary and good, and so on to books not read that were not perceived as literary and perceived as bad. We estimated that we would have enough information to make meaningful observations if each category contained fifty books. Based on this assumption, we went in search of four hundred books.

we included both novels originally written in Dutch and novels translated into Dutch. For comparable reasons we also did not want to limit ourselves to the 'literary novel'. If we wanted to find out what 'literature' and literary quality are, we would have to be able to identify the differences between fiction that does and does not get high scores for literary quality. That is why we also paid a lot of attention to novels that publishers labelled differently than 'Literary novel', such as thrillers and romance, to name only the most well-known. To make our selection replicable we based it on the genre label the Dutch publishers attributed to the books.⁵ We did not include the label in the list of the books as shown to the respondents nor did we show book covers. Publishers often select a book cover with a particular category and audience in mind, to influence buyers and readers. As researchers such as Dixon et al. (2015) have shown, readers can often tell from the cover images what they may expect from a book. To minimize the influence on the participants in The National Reader Survey we only presented the author's name and the book title.

To ensure that we would select books read by as many people as possible – otherwise we risked not gathering enough opinions to make meaningful observations – we decided to compile the list based on the best-selling titles in bookstores and the most borrowed titles in public libraries in the three years preceding The National Reader Survey.⁶ We completed the list with the three book week gifts from 2010, 2011, and 2012, the free book that buyers receive from their bookseller when they buy a book during the yearly 'book week' organized by the Foundation 'Collective Promotion for the Dutch Book' (CPNB). The book week gifts are not included in sales lists, but because their circulation figures are considerably higher than those of most other books – several hundreds of thousands as opposed to

5 In 2013, Dutch publishers made use of NUR, which stands for 'Nederlandse Uniforme Rubrieksindeling' ('Dutch Uniform Classification'). The code consists of three digits and indicates the main subject (e.g. detective, adventure novel), the form (paperback, comics general), or the target group (literary fiction general, fiction 15+) of a book. Booksellers use the NUR code to decide where to place a book in their shop: on the table with new suspense novels or literary fiction, or on the shelves with e.g. poetry or books for young adults. In online bookshops the label associated with a specific NUR code is used to facilitate finding books with a similar content, form or target group. For our research into what constitutes literary quality we focused on NUR code 300, Literary Fiction General, and selected all longer fiction from this category. We added subcategory 285, Fiction 15+ (also known as Young Adult Fiction) from NUR code 200, Children's books.

6 For the selection we used the Top 100 lists of CPNB (the Foundation 'Collective Promotion for the Dutch Book') for the years 2010, 2011, and 2012, and the lending figures of the Stichting Leenrecht (Foundation for Lending Rights) for July 2009–June 2010 and for the same period in 2010–2011 and 2011–2012. Significantly more fiction was bought from bookstores than borrowed from public libraries during 2009–2012 and many of the books that were borrowed often were also on the best-selling list. We addressed this by first selecting 375 titles from the top of the list of best-selling books, supplementing these with 22 titles that were not included in the list of most borrowed books.

several thousands to tens of thousands of other books – their inclusion on our list was justified. Of course, we do not know whether all the copies given away by bookshops were actually read but the same is true for all the other books that were bought or borrowed. Because the survey opened only days before the start of the 2013 Dutch Book Week, we added that year's book week gift as a bonus so the final list consisted of 401 books.

In the course of our research we discovered that our list has a few titles that had slipped through our selection net. We did collect opinions about these books. I will clearly indicate whether these are included in the analyses I present in this book.⁷

The Readers

We asked respondents to indicate not only their age and part of their postal code, but also their gender and their level of education. It may be useful to note that the Dutch survey did not include questions about ethnicity and about class, but that both of these were added to the Big Book Review running in the United Kingdom at the time of writing this book.

When asking participants for their gender we invited them to select one of only three options: female, male, or no information – we are thoroughly aware of the fact that gender is not a binary concept, but a longer list of options would not have been particularly useful for our analysis. Of the Dutch participants who indicated that they read at least one book a year, 72.3 per cent identified as women, 27.1 per cent as men, and 0.6 per cent (ninety people) did not share information. The number of female participants was even higher in the age group of 16–20 years (82 per cent) but slightly lower in the age group of 65 years and older (over 62 per cent). This result agrees with earlier research that has shown that in the Netherlands most readers of fiction were women. That 61.3 per cent of readers aged fifty and over were women is also in line with previous research.

⁷ Occasionally a novel turned out to have been published before 2007, such as Saskia Noort, *De eetclub* (*The Dinner Club*). Several separately published long stories turned out to be from much earlier published collections of short stories, for example Tessa de Loo's *De grote moeder* (**The Great Mother*, the asterisk indicates the book is not available in English) from her 1983 debut collection *De meisjes van de suikerwerkfabriek* (**The Girls from the Confectionery Factory*). And a number of books published as novels turned out to be short story collections. Some readers may also wonder why the list includes books such as *Reizen zonder John* (*In America: Travels with John Steinbeck*) by Geert Mak and *Baltische zielen* (**Baltic Souls*) by Jan Brokken, which many readers would rather call (literary) non-fiction. The publishing house where these books appeared gave them the label literary novel, and that's how they ended up on our list. *Baltische zielen* by Jan Brokken also contains various contributions and could therefore have been regarded as a collection, so that in retrospect the book should not have been admitted to our list.

More than 70 per cent of the respondents indicated that they attained higher vocational education (37.6 per cent) or university education (34.2 per cent). This is 69.4 per cent for women and 78.6 per cent for men. Of the women 27 per cent only completed secondary education while for the men this is 18.1 per cent. The percentages of women and men with primary education as their highest level of education were not that far apart: 2 per cent of women and 1.5 per cent of men selected this option. We did not receive any indication of education level for 1.6 per cent of the women and 1.8 per cent of the men.

Participants represented all regions of the Netherlands. For about 14 per cent the place of residence is not known. Furthermore, a few hundred participants came from outside the Netherlands, mostly from Belgium. They either did not enter a postal code or they input a postal code identical to the first part of a Dutch postal code.⁸ Participants from Flanders might therefore have filled in the same postal code as some participants from the Netherlands. This had a very limited influence on the results, however.⁹

The participants of The National Reader Survey mainly represented readers with a high level of education and as such were not representative of the entire Dutch population. We know from the annual Intomart GfK survey on reading behaviour in the Netherlands in 2012, 2013, and 2014, for example, that around 83 per cent of the Dutch population read at least one book a year.¹⁰ For the 13,784 people who took part in the National Reader Survey, the figure was 98.3 per cent. More details can be found in Figure 2.1, which shows the results of the question 'How many books do you read on average per year?' On the left we see that 6100 participants filled in a number from 1 to 20, almost 3800 participants read 21 to 40 books per year, and as the number of books per year increases, the number of participants rapidly decreases from there. On the right, the figure shows the results for bins of 10: 2350 participants read 1 to 10 books per year, 3700 read 11 to 20, and 2450 read 21 to 30 books, et cetera. In total, 224 participants filled in a number higher than 100, and there were ten people who indicated reading more than 200 books a year, one of them even boasted the amazing number of 700. An interactive version of this graph as well as next graphs (in colour) can be found on the companion website to this book (see Appendix 3).

8 See note 2 to this chapter.

9 We also know the survey attracted participants with a first language other than Dutch, for example from this motivation – I quote the exact text as written by the respondent: 'Sorry my Dutch is not good enough to write an answer to this question. (Sorry mijn Nederlands is niet goed genoeg voor deze vraag)' (ID_5767). The novel this participant had rated was a translation. Most of the times it is not possible to ascertain whether respondents rating translations had read the rated novel in Dutch or in its original language, but we do not consider this to be a problem.

10 See <https://www.leesmonitor.nu/nl/wie-leren-er>, with references to more detailed reports there.

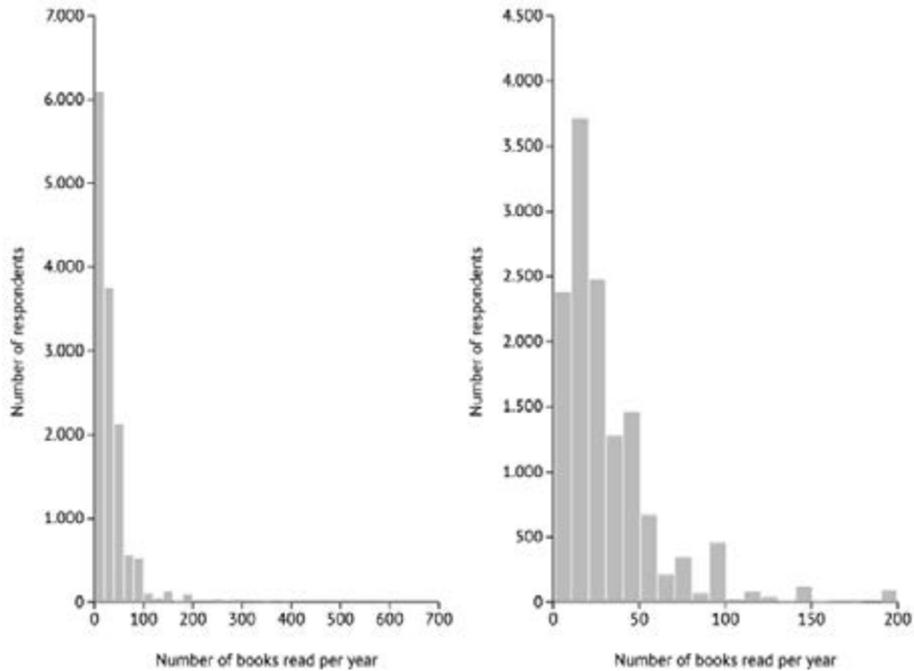


Figure 2.1: Number of books respondents indicated to have read per year

A small number of participants indicated reading less than one book a year. Several hundred also read only non-fiction. We were especially interested in those reading fiction. The total number of fiction-reading participants was 13,332 and they read an average of 28.5 books per year, which amounts to a book about every fortnight and perhaps a few extra during the holidays. Another study from early 2013 found that residents of the Netherlands read an average of fifteen books a year all the way through (that study did not count books that readers did not finish).¹¹ This confirms that the National Reader Survey attracted a lot of frequent readers.

To find out whether the respondents to The National Reader Survey may be distinguished into different types of readers I collaborated with the American literary theorist and statistician Allen Riddell. Many theories address how readers read, but most of these are difficult to test empirically. This is the case, for example, for the four modes of reading proposed in 2008 by literary theorist Rita Felski. She summarized each of these in one word: recognition (to get to know oneself better),

¹¹ See <https://www.kvbboekwerk.nl/app/uploads/2017/09/2013-smb-gfk-meting-23-regulier.pdf>, accessed through <https://www.kvbboekwerk.nl/consumentenonderzoek/consumentenonderzoek>.

enchantment (to have an aesthetic experience, looking for beauty), knowledge (to gain more insight into the society one lives in), and finally shock (to be shocked and confused). The number of questions we would have had to add to The National Reader Survey to find out if our readers may be characterized as reading according to one or more of these modes and thus testing this theory was far too large, and the same was the case for most of the other theories we knew of. One theory that we thought we would be able to test in our survey was put forward by German literary scholars Renate Von Heydebrand and Simone Winko, who suggested that readers take on roughly two main roles: in the one their main aim is an aesthetic experience and in the other it is identification, relaxation, and/or knowledge of other worlds. They hypothesized there are readers who primarily assume one or the other role but that some readers may also switch roles. The reader role in which the primary aim is an aesthetic experience is described by them as 'distanced'; this role is characteristic of readers who usually choose books to which high literary quality is attributed. The other reading role, seeking identification, relaxation, or knowledge of other worlds, they described as 'identifying' (Von Heydebrand and Winko 102–103).

We used the data gathered from 501 participants who completed the survey in September 2013 as members of a hired representative panel of Dutch citizens. Among the sixteen statements presented to the participants, three were phrased in accordance with each of the two reader roles described by Von Heydebrand and Winko (Riddell and Van Dalen-Oskam 5). For the identifying reader, they are:

I read novels to discover new worlds and unknown ages.
 I like novels which are based on true stories.
 I want to be swept away by a novel.

And for the distanced reader:

I like to explore the deeper layers of a novel.
 I like to think about the plot structure of a novel.
 I read novels in order to be challenged intellectually.

Participants indicated to what extent they agreed or disagreed with the statements. We analysed the answers to the six statements and compared them to the answer to another question: 'On a scale of 1–7, where 1 means a non-literary reader and 7 means a very literary reader, where would you place yourself as a reader?'

We found that 84 per cent of all panel members strongly agreed with the statement 'I want to be swept away by a novel'. The statement that was embraced the least – by only 30 per cent – was: 'I read novels in order to be challenged intellectually'. We

then investigated whether the answers to different statements were related. In other words, can we predict someone's response to Statement X when we know what their response was to Statement Y? It turned out that readers who mostly agreed with the statements for the distanced, aesthetic reader role that is usually associated with high literary quality, also mostly agreed with the statements for the identifying reader role. So, if someone agrees with 'I like to explore the deeper layers of a novel', it is very likely that they also agree with 'I like novels which are based on true stories'. However, the reverse is not true. So, for participants who agree with the statement 'I read novels to discover new worlds and unknown ages', we cannot predict with certainty that they also agree with 'I like to think about the plot structure of a novel'.

The conclusion for this representative panel of 501 participants is that the distanced readers (who usually highly value literary quality) very often also were identifying readers (looking for entertainment), but that identifying readers by no means always also were distanced readers. Put another way: actually, most of these readers wanted to be carried away by a novel, so also those who were looking for an aesthetic experience or an intellectual challenge.

If indeed there are only two roles readers may take on, it seems that the distanced role seeking aesthetic satisfaction is complementary to the identifying role, so a kind of extra role. From that perspective, then, many identifying readers have only one role – and so cannot 'switch roles'. Distanced readers do have both roles and could, in principle, switch. In that case, the next question is whether they do switch roles and whether this influences their ratings or not. Are the ratings of distanced readers in their 'identifying role' comparable with the ratings of readers who only read with identifying aims? Or do they give different scores for literary quality to the books they have read than readers who read purely for pleasure? Unfortunately, we were unable to answer this question. The representative panel was too small and even among all of the 13,784 participants in The National Reader Survey we did not find enough readers who only read for identification to test this.

Most readers have ticked as read only a few or, at most, a few dozen of the 401 books on our list. Figure 2.2 (left) shows that 3822 participants read 1 to 10 books out of the 401, 2976 people read 11 to 20, and so on. Figure 2.2 (right) shows how many books participants rated, i.e. how many readers kept asking for another set of seven titles which they wanted to give their opinion about. A total of 6142 participants scored zero to ten books on literary and general quality, 4465 did so for eleven to twenty books. Remarkably, fourteen participants rated one hundred or more books, with a maximum of 149.

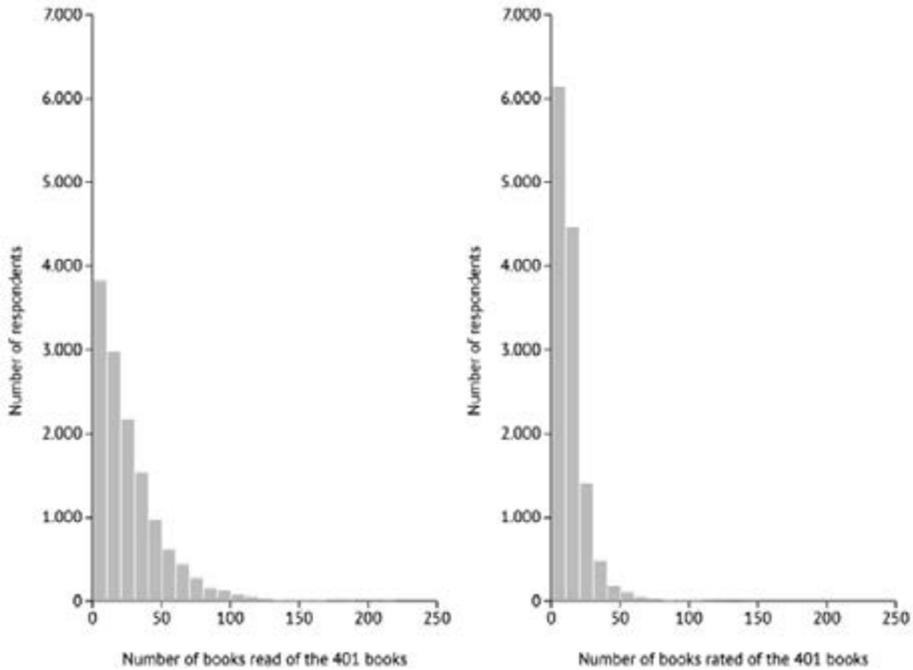


Figure 2.2: Number of books read (left) and number of books rated (right)

Categories of Books

To get to know a little more about the reading patterns of the respondents we compared their engagement with the different categories of novels in the corpus. For our analysis we divided the books into four different categories – which we did not show to the participants to avoid influencing their rating.¹² The four categories were Literary novel, Suspense, Romance, and Other. Of the 401 novels 249 were translations and 152 were originally written in Dutch. A total of 186 books of the 401 were categorized as thrillers or literary thrillers – we brought them together under the label of Suspense. Of these 186, 58 were originally written in Dutch and 128 were translated. The origin of the Suspense novels that sold best in the Netherlands and were borrowed most from public libraries is remarkably homogeneous: they are of Dutch origin or have been translated from English or from

¹² We used the NUR codes (see note 5 to this chapter) publishers attributed to the books and whenever publishers used different labels for different print runs of the same book additionally inspected the labels used by NBD Biblion, an organization supporting Dutch libraries, as a kind of second opinion.

one of the Scandinavian languages. One thriller (by Deon Meyer) was translated from Afrikaans. Although the entire category contains more male authors than female authors, 43 versus 29, the number of titles differs less: 103 versus 83. The fact that female authors account for 83 titles is mainly due to the popularity of the Dutch (literary) thriller. I should add that for the corpus used in the Dutch National Reader Survey we selected all books of the same author that passed our selection criteria. This resulted in, for example, nine novels by Suzanne Vermeer (which for most of the titles is the pen name of a male author, by the way), eleven by David Baldacci, and thirteen by Loes den Hollander. When we started work on the selection of the English-language corpus for the replication study in the United Kingdom, we decided that we wanted to cap the total number of titles per author at three, so as to present as wide a set of authors as possible but without completely ignoring their popularity.

The Romance category contained 41 titles. These included five novels written in Dutch and 36 that had been translated, all from the English and all by female authors. The Dutch Romance novels were also authored by women except for one, which was co-authored by a mixed-gender duo.

The corpus contained a total of 147 titles in the category Literary novel – in Dutch usually labelled as ‘*roman*’ (roman); the equivalent in English will mostly be ‘fiction’ or ‘general fiction’. Of those 147, 74 were translated and 73 were originally Dutch. The translations came from German, English, French, Italian, Japanese, Portuguese, Spanish, and Swedish. As to language of origin the category is more diverse than that of Suspense but still very much Western and white.

The category Other comprised 27 books, of which sixteen were original Dutch and eleven had been translated into Dutch. The translations were all from English. Of these books, thirteen belonged to a genre that is represented in our list with only a few titles.¹³ In this category we also placed the eleven books that turned out not to be longer fiction, but a collection of stories or literary non-fiction. Finally, we also placed the Dutch translations of the immensely popular books by E.L. James, *Vijftig tinten grijs* (*Fifty Shades of Grey*), *Vijftig tinten donkerder* (*Fifty Shades Darker*), and *Vijftig tinten vrij* (*Fifty Shades Freed*) in this category. The Dutch publisher labelled them as Literary novels, but this categorization seemed counter-intuitive to us. Keeping them apart from the three main categories made it possible to check with computer software whether the categorization Literary novel was indeed the most appropriate one. I will present an answer to this question in Chapter 3.

Figure 2.3 shows per category how many participants ticked 1-5, 6-10, 11-15, et cetera, books from each of the category as read.

¹³ These are NUR codes 340 Popular Fiction General and 344 Regional and Family Novels, as well as a few titles labelled Fantasy and Young Adult (Children’s Books Fiction 15+). More about NUR in note 5.

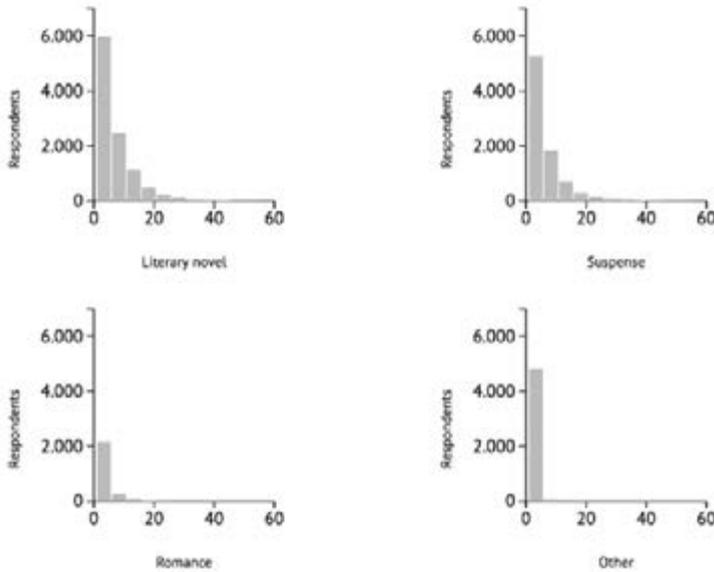


Figure 2.3: Number of books read per category.

Total number of books per category: Literary novel 147, Suspense 186, Romance 41, Other 27.

Winners and Losers

The novel most read by the respondents was *Haar naam was Sarah* (*Sarah's Key*) by Tatiana de Rosnay: almost half of all participants indicated to have read it. In the second place we find the Dutch novel *Het diner* (published in English as *The Dinner*) written by Herman Koch, and his next novel also appeared in the top 10 read (Table 2.1).

Table 2.1: Top 10 most read of the 401 books

Number of readers	Author, Title (Title of published English version; *not available in English in 2022)
7135	Tatiana de Rosnay, <i>Haar naam was Sarah</i> (<i>Sarah's Key</i>)
6834	Herman Koch, <i>Het diner</i> (<i>The Dinner</i>)
5638	Paolo Giordano, <i>De eenzaamheid van de priemgetallen</i> (<i>The Solitude of Prime Numbers</i>)
5623	Stieg Larsson, <i>Mannen die vrouwen haten</i> (<i>The Girl with the Dragon Tattoo</i>)
5598	Carlos Ruiz Zafón, <i>De schaduw van de wind</i> (<i>The Shadow of the Wind</i>)
5247	Peter Buwalda, <i>Bonita Avenue</i> (<i>Bonita Avenue</i>)
4728	Stieg Larsson, <i>De vrouw die met vuur speelde</i> (<i>The Girl Who Played with Fire</i>)
4629	Saskia Noort, <i>De eetclub</i> (<i>The Dinner Club</i>)
4300	Herman Koch, <i>Zomerhuis met zwembad</i> (<i>Summer House with Swimming Pool</i>)
4297	Khaled Hosseini, <i>Duizend schitterende zonnen</i> (<i>A Thousand Splendid Suns</i>)

We had hoped to gather at least fifty opinions about all 401 books, because we feared that with fewer opinions, individual readers would have too great an influence on the average score and coincidence would play too big a role in the results. A total of sixteen novels did not reach the threshold of fifty reader opinions. In Appendix 2, which lists all 401 books ranked from highest to lowest score for literary quality, these sixteen are marked with an asterisk.¹⁴

The top 10 most literary books were all written by men (Table 2.2). At the very top we find *Alsof het voorbij is* by Julian Barnes, a translation from the English *The Sense of an Ending*. The English original received the highly prestigious literary Man Booker Prize in October 2011. More about this novel and its Dutch readers in Chapter 5, Literary Novels Written by Men. What is also interesting is that the top 10 includes six Dutch-language novels of which three are by Flemish authors. The three Flemish novels score even higher (6.4, 6.5 and 6.6) than the three Dutch novels (6.2 and twice 6.3).

Table 2.2: Top 10 most literary of the 401 books

Mean score	Author, Title (Title of published English version; *not available in English in 2022)
6.6	Julian Barnes, <i>Alsof het voorbij is</i> (<i>The Sense of an Ending</i>)
6.6	Erwin Mortier, <i>Godenslaap</i> (<i>While the Gods Were Sleeping</i>)
6.5	Erwin Mortier, <i>Gestameld liedboek</i> (<i>Stammered Songbook</i>)
6.4	Michel Houellebecq, <i>De kaart en het gebied</i> (<i>The Map and the Territory</i>)
6.4	Tom Lanoye, <i>Sprakeloos</i> (<i>Speechless</i>)
6.3	Haruki Murakami, <i>Norwegian Wood</i> (<i>Norwegian Wood</i>)
6.3	A.F.Th. van der Heijden, <i>Tonio</i> (* <i>Tonio</i>)
6.3	Stephan Enter, <i>Grip</i> (* <i>Grip</i>)
6.2	J. Bernlef, <i>Geleende levens</i> (* <i>Borrowed Lives</i>)
6.2	Umberto Eco, <i>De begraafplaats van Praag</i> (<i>The Prague Cemetery</i>)

The highest ranked novel written by a female author is *Het grote huis* (from English *Great House*) by Nicole Krauss, with a mean score of 6.0, ranked 32. The first Dutch

14 What is striking is that all five novels with NUR codes 340 Popular Fiction General and 344 Regional and Family Novels, categorized as Other are amongst these. We categorized them as such because this limited number of books from this NUR code cannot help us much further in determining what exactly literary quality is. Moreover, there is no variation in the literary appreciation: they were all experienced as not very literary by the participants in the National Reader Survey. In theory it is possible that there are also regional and family novels that are considered to be highly literary, but they are not in our research corpus, or – which I deem more probable – they have been given a different genre label (Literary novel) by their publishers. More about such genre issues in later chapters. It is also noteworthy that these five novels are among the 22 titles that did not appear on the list of the best-selling books from the years 2010–2012, but during that period were borrowed exceptionally often from public libraries. For the rest, the list of books with less than fifty readers opinions is a mixed bag; it also includes a couple of novels translated from English, both from the categories of Literary novel, Romance, and Suspense.

novel written by a woman ranks 36: Margriet de Moor, *De schilder en het meisje* (**The Painter and the Girl*, not available in English, mean score 5.9). Both these novels will receive due attention in Chapter 4, Literary Novels Written by Women.

A closer look at the bottom ten, however, is also extremely interesting (Table 2.3). At the very bottom, E.L. James' *Vijftig tinten grijs* (from English *Fifty Shades of Grey*) has a score of 2.1. The next two parts of this trilogy both scored 2.6 and are at the bottom of the list at places 12 and 14. The hugely popular *Fifty Shades of Grey* trilogy is a phenomenon that deserves special attention. In Chapter 3, Romance, Suspense, and Translations, I will discuss these books in detail and will also address the difference in score between Volume 1 and Volumes 2 and 3.

Table 2.3: Top 10 least literary of the 401 books

Mean score	Author, Title (Title of published English version; *not available in English in 2022)
2.6	Jill Mansell, <i>Drie is te veel</i> (<i>Two's Company</i>)
2.6	Katie Fforde, <i>Trouwplannen</i> (<i>Wedding Season</i>)
2.5	Lauren Weisberger, <i>Champagne in Chateau Marmont</i> (<i>Last Night at Chateau Marmont</i>)
2.5	Jill Mansell, <i>Versier me dan</i> (<i>Take a Chance on Me</i>)
2.5	Sophie Kinsella, <i>Ken je me nog?</i> (<i>Remember me?</i>)
2.5	Sophie Kinsella, <i>Mini shopaholic</i> (<i>Mini Shopaholic</i>)
2.5	Lauren Weisberger, <i>Chanel chic</i> (<i>Chasing Harry Winston</i>)
2.4	Sophie Kinsella, <i>Mag ik je nummer even?</i> (<i>I've Got Your Number</i>)
2.4	Sophie Kinsella, <i>Shopaholic en baby</i> (<i>Shopaholic and Baby</i>)
2.1	E.L. James, <i>Vijftig tinten grijs</i> (<i>Fifty Shades of Grey</i>)

All ten novels at the bottom of the literary ranking were written by female authors and, apart from *Vijftig tinten grijs* (*Fifty Shades of Grey*), they all belong to the category of Romance. Moreover, all ten are translations from English. At place 16 from the bottom, we find the first book by a male author: *Een dief in de nacht* (**A Thief in the Night*) by Appie Baantjer and Simon de Waal, which received an average score of 2.7.

That literary quality and general quality indeed are two different things for the participants in The National Reader Survey becomes apparent when the list is sorted from best to worst. *Vijftig tinten grijs* (*Fifty Shades of Grey*) is still at the very bottom but the average score of 3.5 for general quality is, on a scale of 7, already less serious than the 2.1 for literary quality. *Vijftig tinten donkerder* (*Fifty Shades Darker*) is now also in the last ten; the third volume can be found at place 13 from the bottom. *Vijftig tinten vrij* (*Fifty Shades Freed*) scored an average of 4.6 and again we see a striking difference between the first volume on the one hand and the second and third volumes on the other. The other eight at the bottom are completely different from the bottom of the list for literary quality. The Romance novels that received such low ratings for literary quality interestingly are not rated as the worst

novels one could read. Next to the two *Fifty shades* volumes, only one of the ten is a translation: Elizabeth Gilbert, *Eten, bidden, beminnen* (*Eat, Pray, Love*), which scored a 4.6 for general quality. What is also striking is that the other seven books are originally Dutch novels, written by both female and male authors; four of these are from the category Literary novel, two from Suspense, and one from Romance.

Readers did indeed make a difference in their judgements between literary quality and general quality. This becomes even clearer when we compare the average scores for all 401 novels. Figure 2.4 shows how many books have an average score ranging from 0–1, 1–2, et cetera, to 6–7.¹⁵ The lowest average score for general quality, i.e. how good readers thought the book was, is 4. The lowest average score for literary quality is 2. So we see a broader range for literary quality: from 2 to 6. If we compare the mean scores for literary quality and general quality for all books, we see that two thirds of the titles labelled by the publishers as Literary novel score lower on literary quality than on general quality, while one third score the same or slightly higher. Books from the other two major genres in our corpus, Romance and Suspense, without exception score higher on general quality than on literary quality. The biggest difference in score for Romance novels is 3.1, for *Mag ik je nummer even?* (*I've Got Your Number*) by Sophie Kinsella, with a 5.5 for general quality and 2.4 for literary quality. In the Suspense category, the biggest difference is 2.7, for *61 Uur* (*61 Hours*) by Lee Child (overall quality: 5.9, versus literary quality: 3.2).

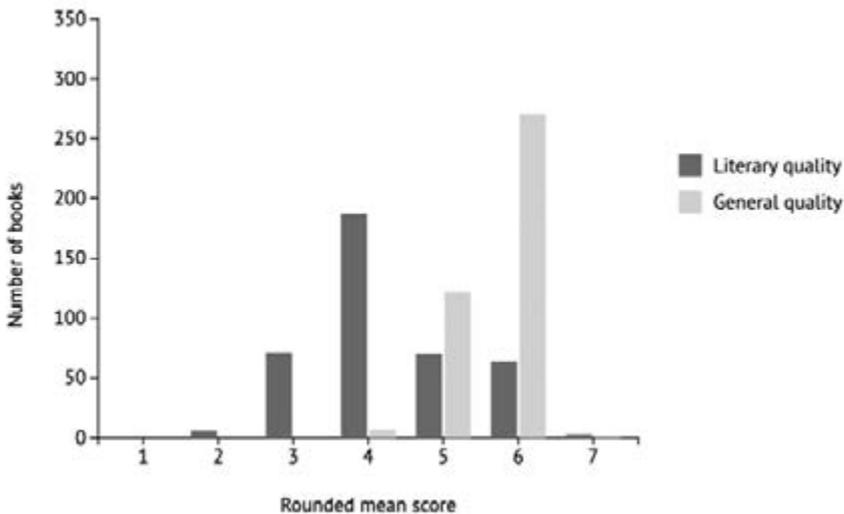


Figure 2.4: Mean scores for all 401 books

¹⁵ In other words, Fig. 2.4 presents the aggregated number of books having a mean score rounded to the nearest whole number.

Not reading a book can also reflect an interesting opinion. Participants in The National Reader Survey could indicate whether there were any books on the list that they had not read but did have an opinion about. Participants could then rate a maximum of seven of these books. The lowest number of 'not read, but have an opinion' for a book is 19, the highest 4310. As a reminder, a total of 13,784 people completed the survey. The top 10 of books participants most often have an opinion about is presented in Table 2.4.

Table 2.4: Top 10 books participants had not read but had an opinion about

Number of participants	Author, Title (Title of published English version; *not available in English in 2022)
4310	E.L. James, <i>Vijftig tinten grijs</i> (<i>Fifty Shades of Grey</i>)
2139	E.L. James, <i>Vijftig tinten donkerder</i> (<i>Fifty Shades Darker</i>)
1849	A.F.Th. van der Heijden, <i>Tonio</i> (* <i>Tonio</i>)
1728	Kluun, <i>Haantjes</i> (* <i>Alpha Males</i>)
1716	E.L. James, <i>Vijftig tinten vrij</i> (<i>Fifty Shades Freed</i>)
1019	Stieg Larsson, <i>Mannen die vrouwen haten</i> (<i>The Girl with the Dragon Tattoo</i>)
983	Jonas Jonasson, <i>De 100-jarige man die uit het raam klom en verdween</i> (<i>The 100-Year-Old Man Who Climbed Out the Window and Disappeared</i>)
976	Tatiana de Rosnay, <i>Haar naam was Sarah</i> (<i>Sarah's Key</i>)
976	Herman Koch, <i>Zomerhuis met zwembad</i> (<i>Summer House with Swimming Pool</i>)
951	Saskia Noort, <i>Koorts</i> (* <i>Fever</i>)

The list in Table 2.4 probably primarily highlights which books were most visible in bookshops, in newspapers, on television and elsewhere during the time preceding our survey in 2013 and were most often discussed and talked about. The top 10 includes the Dutch translation of all three volumes of E.L. James' *Fifty Shades* trilogy, and two Dutch novels that appeared almost simultaneously in early 2011: *Zomerhuis met zwembad* (*Summer House with Swimming Pool*) by Herman Koch and *Haantjes* (**Alpha Males*) by Kluun (a one-word penname of the author publishing in English as Ray Kluun). This almost simultaneous publication received a lot of media attention for months on end. In Chapter 5, *Literary Novels Written by Men*, I will analyse the head-to-head literary battle this resulted in.

We noted an interesting trend when we compared the mean scores for all 401 books by readers and non-readers. Novels ending up at the bottom of the literary rankings were usually judged even more harshly for their literary quality by non-readers than by readers. The books at the top of the literary rankings, however, received roughly the same score from non-readers as from readers and sometimes the scores were even a little higher. This firmly illustrates that the participants in

The National Reader Survey did have an idea of the literary value of books they had not read. The implication is that not only the books themselves played a role in the assessment but also their reputation or that of their author. In short: there were indeed certain value conventions and these were apparently recognizable to readers without even having to read the book.

The four main categories of books also received clearly different mean scores (Table 2.5). The average score on general quality is 5.81 (on a scale of 7) for all 147 novels in the category Literary novel, 5.66 for all 186 novels in the category Suspense. The 27 titles labelled by us as Other scored an average of 5.59 and the 41 in Romance received an average of 5.27 from readers, which is the lowest average score per category. So, the differences in overall quality (how good is the book?) between the four main categories were not much more than half a point. The averages were calculated on the basis of all reader opinions together. Taking the average score per book, the numbers were different but the trend remained the same. This also means that the corpus would be of little help if we wanted to find out what distinguishes a good book from a bad one: it simply did not contain enough bad books. Luckily, for literary quality the differences were considerably larger. Romance scored lowest, with an average of 2.95, and the category of the Literary novel received the highest mean score, with 5.44. Suspense ended up in between these two categories with 3.94. Other is just above Suspense with 4.07.

Table 2.5: Mean scores for literary quality and general quality per category

Category	Number of books	Literary quality	General quality
Literary Novel	147	5.44	5.81
Suspense	186	3.94	5.66
Romance	41	2.95	5.27
Other	27	4.07	5.59

It is also informative to look at the distribution of rounded mean ratings for literary quality across all categories (Figure 2.5). We see, for example, that the majority of Literary novels received an average score that was rounded to 5 or 6 for literary quality, and most Suspense had a mean rate rounded to 4). Both categories most often received a rounded to 6 for general quality (Figure 2.6). In other words, most Literary novels had a higher score on literary quality than most other books. Apparently, a suspenseful book was hardly ever highly literary. Romance had even less chance of being rated as highly literary. There was an overlap, but the ranking of the genres was clear. Literary novels were at the top, and Romance at the bottom of the genre hierarchy.

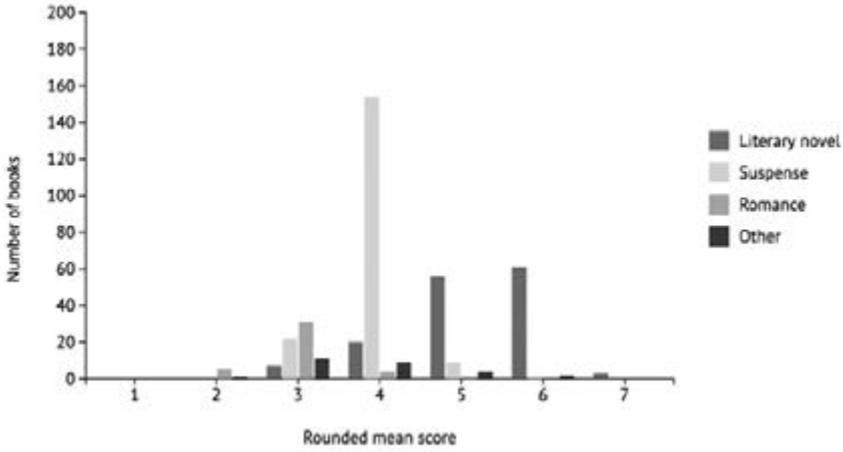


Figure 2.5: For all 401 books the mean score for literary quality per category

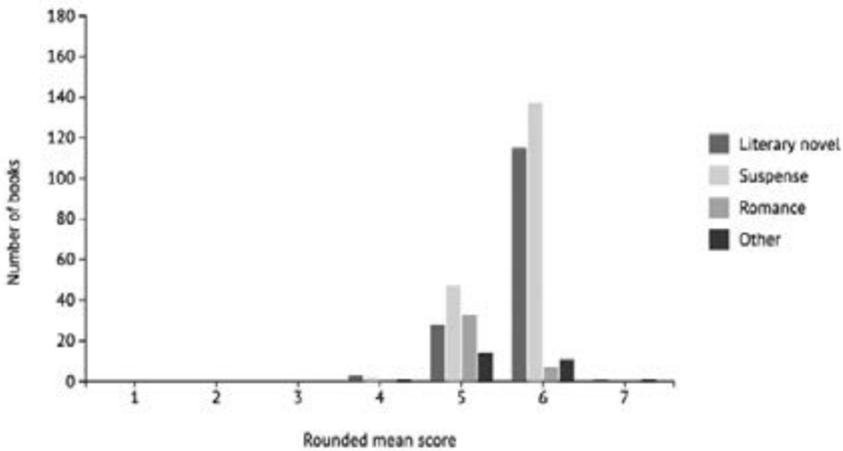


Figure 2.6: For all 401 books the mean score for general quality per category

An additional feature we considered is whether a book was written in Dutch or was translated into Dutch from another language. Taking all reader opinions on all 401 books into account, we found that their average score for literary quality was 4.67 (on a scale of 7 being the highest). The 152 original Dutch novels received an average of 4.95 from the participants in The National Reader Survey, which was slightly higher than the overall average, while the 249 translated novels received an average of 4.47, slightly lower than the average of 4.67. When it comes to literary quality, translations were thus rated slightly lower than originals, and this difference is statistically significant.

The possible influence of the translation of a novel on the reader's judgement is further explored in Chapter 3, Romance, Suspense, and Translations.

Zooming in on individual books on the list also helps to find out more about the diversity in readers' ratings. The novel that was most read, *Haar naam was Sarah* (*Sarah's Key*) by Tatiana de Rosnay, was rated for general and literary quality by 1117 participants. Not all of the 7135 readers who had indicated to have read this novel were presented with the title for assessment; our survey software ensured that as many different titles as possible were rated by the participants. As a reminder: score 1 was the lowest score and stood for 'definitely not literary' and 7 was the highest, 'highly literary'. Participants could only enter whole numbers, but the averages we calculated obviously have decimal points. In the eye of the readers, *Haar naam was Sarah* (*Sarah's Key*) scored an average of 4.7 for literary quality and 5.7 for general quality (how good the book was considered to be). Figure 2.7 shows that for this novel, the most commonly awarded score for general quality is a 7 and for literary quality a 6. For the sake of completeness: the non-readers with an opinion about this book gave it slightly lower scores: an average of 4.0 for literary quality and 4.4 for general quality.

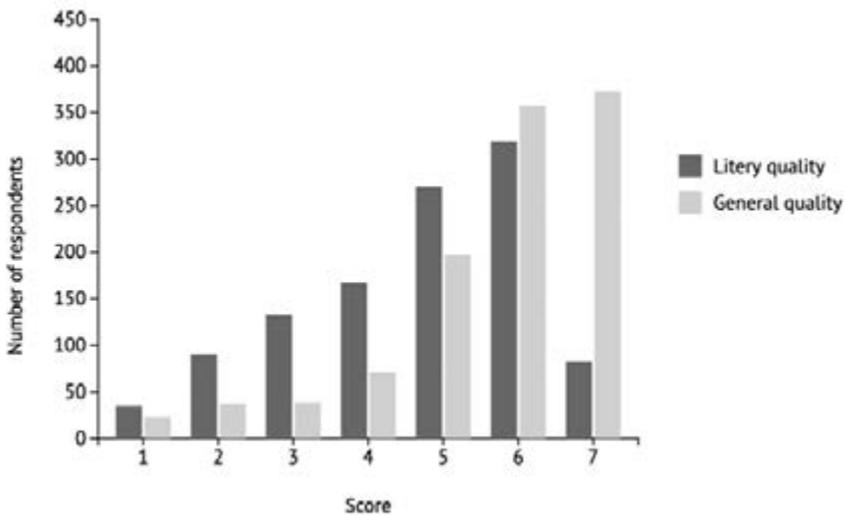


Figure 2.7: Ratings of Tatiana de Rosnay, *Haar naam was Sarah* (*Sarah's Key*)

Motivations

The National Reader Survey not only yielded ratings of the 401 books on the list, but for one book they rated participants were also asked to motivate their score for

literary quality in a free text field. The survey software selected a book that ranked relatively high or low on the literary quality scale compared to the other books rated by the same person. Around 11,951 of the 13,784 respondents made use of this opportunity. The answers ranged from one word (often 'no') up to several sentences. With just a couple of exceptions these motivations were written in Dutch. I will present them in my own English translation; the respondent ID added between brackets after my translation may lead the interested scholar to the original Dutch text in the litRiddle package for R (see Appendix 3).

By way of introduction, I present a selection of motivations about *Haar naam was Sarah* (*Sarah's Key*) by Tatiana de Rosnay. One of the participants justified the score 7, 'highly literary', as follows: '*Haar naam was Sarah* has several storylines that adequately come together at the end. Furthermore, it is written in such a way that the person reading it feels emotionally involved. Also, the author's choice of words is "more difficult" compared to, for example, books by Collins and Rowling' (ID_7261). Another reader wrote: 'Several storylines run through each other, which gives the book extra power. The plot is clear, and you can read on without having to refer back to earlier parts of the book. There are beautiful sentences in it' (ID_8559). A score 6, 'literary', was motivated as follows: 'The story, the construction, the suspense, the human condition, beautifully rendered' (ID_1693). The next participant found the novel 'tending towards not literary' (score 3): 'I find the style extremely annoying. In fact, it is a glorified kitchen maid's novel' (ID_10070). Score 2, 'not literary', was justified by a respondent as follows: 'It is mainly a compelling story, but it is not well written. Nothing is left to the imagination of the reader' (ID_8092). And 'definitely not literary', the lowest score of all, was given to the novel by the following reader: 'A cliché story, structured according to a fixed pattern, everyday language, nothing surprising. All emotions overworked'. (ID_9302).

On average, the motivations consist of 15.6 words. All motivations together form a quite substantial research corpus that can provide us with a lot of useful information. Printed they take up 550 pages and it would take weeks to categorize or label them in a useful way for further analysis. However, with a few simple actions in a simple computer programme, it is possible to gain insight into the most important trends very quickly. I used the concordance software AntConc, developed by Laurence Anthony (2019), to get an overview of all the words used and their frequencies and in this way I selected which features to further zoom in on. The programme indicated that all motivations together contained 186,530 words (tokens), divided into a total of 10,728 different words (types). More than half of these types occurred only once, a total of 5792.

A very simple first step is to explore the words used most frequently in all motivations. Amongst the hundred most frequent words were nouns such as (I

have translated the Dutch words into English): *story, style, language, depth, layers, sentences, characters, way, plot, structure, content, theme*, and adjectives such as *good, literary, beautiful*, and *suspenseful*. Most prominent were motivations in which *style* played an important role, both positively and negatively.

I did not care for the **writing style** (ID_9745).

Because of the **style**, the construction of the story, the terms used, and especially the subject matter, described from an unusual point of view (ID_8476).

Very often, participants also mentioned the absence or presence of *deeper layers* as the reason for their score for literary quality.

Book has a lot of depth and multiple **layers** (ID_6542).

It is suspenseful, the storyline is well crafted, but in literary books I expect an **extra layer** (ID_9532).

Different arguments were regularly combined. For instance, reference to the *characters* in the story was accompanied by a mention of the style and deeper layers.

Beautiful sentences and word choice. **Characters** studied in depth. Double layer (ID_2409).

It's just a fun book, purely for relaxation. The **characters** don't develop, there are no beautiful sentences, you don't have to search for the deeper meaning, etc. (ID_9230).

The same applies when respondents mentioned the *story* and the *structure*, the *plot* – usually style and deeper layers are also addressed.

Complicated **structure**, beautiful sentences, nice use of words, cultural theme, great attention to atmosphere (ID_12253).

Both the **story** and the **construction of the story** are special, the book has deeper layers. It also has intertextuality. Furthermore, I would call the writing style literary (ID_11713).

Some participants explicitly wondered what literature or literary quality actually is, or even went in search of a definition – which we deliberately did not provide because we did not want to influence the respondents.

Literary, of course, is a tricky word. What is **literature** but what those who know about literature think is literature? Nevertheless, I think this is literature. It is well constructed, and the characters are well developed (ID_3837).

From Wikipedia I get: In general, **literature** is understood to be the entirety of texts considered to have universal value. In my opinion, this is not the case with *Herfstlied* (*Autumn Song). It is purely for entertainment (ID_5903).

Some participants referred to what they learned in high school for their answers.

Had to read it for the reading list. Not a very suspenseful story. There are questions but they didn't stimulate me (ID_8088).

Complexity of the story, character development, writing style. More or less the characteristics I used to learn in high school, although these can be applied to more books than my literature teacher listed (ID_4065).

The association of literature with difficult and boring came up several times.

A book without a good story that does not give pleasure in reading. That has got to be literature (ID_3894).

When I think of literature, I think of a bit of pomposity, with a lot of intertextuality.

This book was so easy to read that it could hardly be literature (ID_3159).

I found it easy to read and sometimes have the feeling that literature has to be difficult to read to be called literary (ID_13077).

Motivations regularly included a reference to the name of an author the respondent attributes more value to than the one in our list of books they were asked to substantiate their rating about. Dutch or Flemish authors were mentioned, such as the well-known (for Dutch readers) literary author Harry Mulisch (1927–2010, not represented on our list), but also foreign authors whose work may have been read in the original language or in translation in the Netherlands.

I think the book is more of a side dish, a real women's book, not suitable for people who like literature. Thinking about literature, I'd rather think of Harry Mulisch: difficult words, fat books that make you think and that's not what I think of this book. Although I liked the book! (ID_6205).

Literary has some depth in its themes and it also has to do with writing style. I consider Sebald and Roth to be literature – without short-changing Nico Dijkshoorn. His observations are sharp and accurate, but there is no great idea behind them – other than to inform the reader about his youth (ID_3832).

Several readers wrote that they did not want to distinguish between what is literary and what not.

No, I think it's nonsense to ask these questions anyway, what difference does it make if something is literary (ID_13629).

I'm not knowledgeable about literature and I'm not interested in the literary value of any book (ID_8402).

On the one hand, because it lacks layers and intertextuality (those are the things that would make me consider a book literary), and on the other hand, since reading Karel van het Reve's Huizinga lecture, I've actually doubted the usefulness of those criteria as arguments for literature – so maybe it's just a feeling (ID_12486).

That their rating for literary quality was 'just a feeling' is what made it difficult for several respondents to explain their opinion.

No, I personally don't have a good definition of literature. It usually is a feeling (ID_12081).

No I can't, it's a feeling (ID_12350).

Contrasting Opinions

In the previous section, a look at the one hundred most frequent words in all the motivations together presented us with a fascinating first view on how the participants in The National Reader Survey motivated their opinions about literary quality. A good way to get more detailed information is to compare the vocabularies of different subsets of the motivations to see where they may differ. For these experiments I again made use of the AntConc software developed by Laurence Anthony (2019). I used AntConc's keyword function to find out which words occur much more often than would be expected in one set of motivations as compared with another. For each of my keyword measurements I used the following values: Log-likelihood (4-term), statistic threshold $p < 0.05$ (+ Bonferroni), effect size measure Dice coefficient.¹⁶ Using the litRiddle package for R and AntConc, my measurements can be easily reproduced, so I will focus on the main results and observations, again presenting English equivalents of the Dutch keywords.

To get an idea of the differences in motivations of a high score for literary quality with those of a low score, I compared motivations for ratings 1, 2, or 3 (low literary quality) with those for ratings 5, 6, and 7 (high literary quality), leaving out 4, the rating exactly in the middle, 'bordering literary and non-literary'. The

¹⁶ More information about keyword analysis and a good example of how it may be used for a stylistic analysis of a literary text can be found in Culpeper 2015.

results showed that motivations for low scores were mainly characterized by words suggesting the lack of something: *no, not, little, nothing*, and *lacks*. They also showed a noticeably higher number of words that indicated a genre, a type of novel: *chick lit, thriller, detective, children's book, fantasy*, which seems to confirm the genre hierarchy in which the literary novel is the apex of literary prestige and genre fiction has a significantly lower literary reputation, as described earlier in this chapter. Keywords for the low scores included qualitative adjectives such as *mediocre, bad, weak, boring*, and the like, but also *delicious, nice, funny, entertaining*, and *romantic*. And what about *simple, ordinary, easy, predictable, superficial, light-hearted*, and *implausible*?

In the motivations for high scores for literary quality, compared to those about low scores, we also saw a great many descriptive adjectives as keywords, but clearly of a different order: *good, wonderful, beautiful, evocative, clever, captivating, clear, poetic, poignant, splendid, excellent, compelling*. What was also striking is that many nouns played a prominent role, often relating to the linguistic form of the book: *composition, construction, perspective, writing style, style, language, theme, storylines, words, sentences*, and so forth. In addition, we noticed other entities besides the protagonist: *reader, human being, mother, writer*. Various keywords have to do with (the passage of) time: *event, history, present, time, past*. More difficult to describe but also of importance were *feelings, empathy, atmosphere, choices, life, world*, and *reality*.

All in all, we can state that the motivations for less literary novels more often pointed to things lacking (such as a certain style and deeper layers), the fact that it was genre fiction, and the accessibility and readability of the work – i.e. the pleasure of reading. The motivations for books considered to be very literary focused on structure, characters, and linguistic properties such as the poetic aspect of the language, and also addressed the historical context. More distant descriptions, one could say, and although feelings and empathy are discussed in the novels, relaxation and enjoyment seem to be hard to find for some readers.

By comparing the motivations of different age groups we might be able to uncover changes over time in the way readers motivate their ideas about literary quality. For this, I again left out the 'middle': I compared the words used by 16-year-olds to 45-year-olds with those used by participants aged 61 and over. The results showed that young people proportionally used more words referring to pleasure and accessibility on the one hand and a challenge on the other. The older ones seemed to write more about aesthetics and admiration. Both groups paid extra attention to certain authors or characters. Among young people, these were *Ammaniti, Harry, Potter, Worthy* (whereby *Harry* usually belongs to *Potter*, and twice to the Dutch author *Mulisch* mentioned earlier in this chapter). Older people talked more often about *Baantjer, Bernlef, Siebelink, Springer*, and *Voskuil*, all Dutch authors. From

other keywords it seemed that the ubiquitous ‘deeper layers’ characterized the younger generations. The older ones wrote more about persons and characters, atmosphere, and content. It was also striking that older people reflected more on translations and translators. Whether or not some of these differences are closely related to changes in the way literature was taught at Dutch secondary schools in recent decades requires further investigation.

It was difficult to confirm any differences between the motivations of respondents with different levels of education. Very few people who took part in The National Reader Survey indicated that they had no education or only primary education, or LBO/VBO/VMBO-kader or vocational/MBO 1 (see the list from which respondents could choose in Appendix 1, The Survey). I grouped their motivations together with those of the respondents with a secondary school education and called the combined group ‘secondary education’. I compared the motivations of this group with those of the respondents with HBO (vocational training) or WO/postgraduate education (university level education) and referred to this group as ‘higher education’. The motivations of the 237 respondents who did not state their education have been left out. People from the ‘secondary education’ group more often mentioned readability and accessibility than those from the other group. In the higher education group, more references to *layering*, *deeper layers*, and *style* were found. The motivations of the group with secondary education seemed to emphasize the content of the story, especially of the Suspense genre. The group ‘higher education’ used more words about the style and form of the novels than about their actual content – more distant. In the group ‘secondary education’ several names were mentioned considerably more often: *Baantjer* (the Dutch author of crime fiction) and *Loes den Hollander* (mainly known as a Dutch thriller author). In the group ‘higher education’, two other names stood out, *Spinoza* (the philosopher) and *Springer* (the Dutch author F. Springer), and a language: *English*, of which many respondents indicated that they would rather read an English original than a Dutch translation. This ties in with the appearance of the noun *translation* in the list of relatively more frequent words in the motivations of the higher educated.

The same method uncovered some differences between the motivations written by women and written by men. The survey yielded a total of 8764 motivations from women and 3109 from men and 77 motivations from participants who checked the ‘no information’ option about their gender. These 77 I left out of this comparison because the number is too low for reliable observations. The women showed a preference for the following adjectives: *deeper*, *fine*, *feel good*, *instructive*, *nice*, *easy*, *beautiful* (Dutch ‘mooi’), *pleasant*, *romantic*, *timeless*, *predictable*, *strange*. Men more often used *beautiful* (Dutch ‘fraai’, a higher register than ‘mooi’), *bright*, *literary*, *original*, *stylistic*. Women more often use nouns such as *chick lit*, *information*,

subject, holiday book. For men, some noun keywords were *composition, memories, main characters, irony, plot, theme*. Women tended to focus somewhat more on reading pleasure and men more on literariness.

However, it also became clear that some differences may be closely related to different reading patterns of female and male participants in regard to the gender of the authors they read. Men more often mentioned the male authors *Houellebecq, Van Kooten, and Mak*. Women more often mentioned *Binchy and Dorrestein*, both female authors. Women also more often mentioned the name *Arthur*, invariably referring to Dutch author Arthur Japin. Both men and women used his surname more or less equally, but women used his first name more often. Furthermore, women more often used the personal pronouns *I, me, you*, and the possessive pronoun *her*, which usually referred to a female author. Men used the personal pronoun *he* relatively more often, which usually referred to a male author.

The survey gathered 4130 motivations written by women about books with a female author, and 4376 written by women about books with a male author – for these calculations I excluded books written by mixed duos and unknown authors. For the male respondents the situation was quite different: we received 599 male-authored motivations about books written by women, and 2436 – four times as many – about novels written by male authors. We noticed a similar discrepancy in the number of books read of the list of 401 books: the male participants read considerably more male-authored novels, whereas the women read almost as many books by female as they did by male authors (Figure 2.8). The same holds for reviewed novels (Figure 2.9).

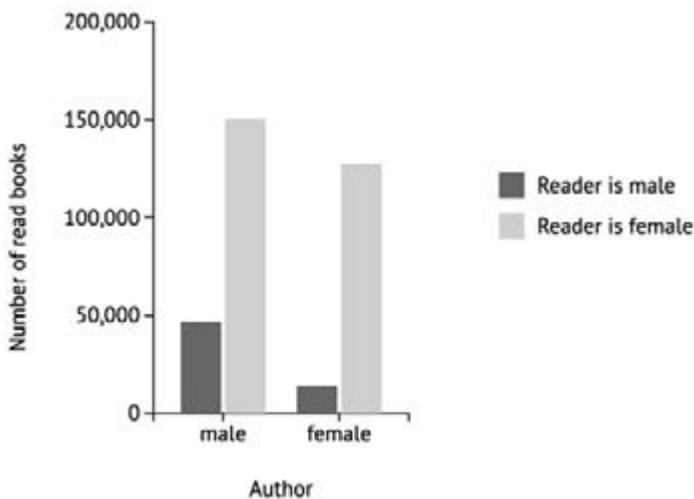


Figure 2.8: Number of books read, by gender author and gender reader

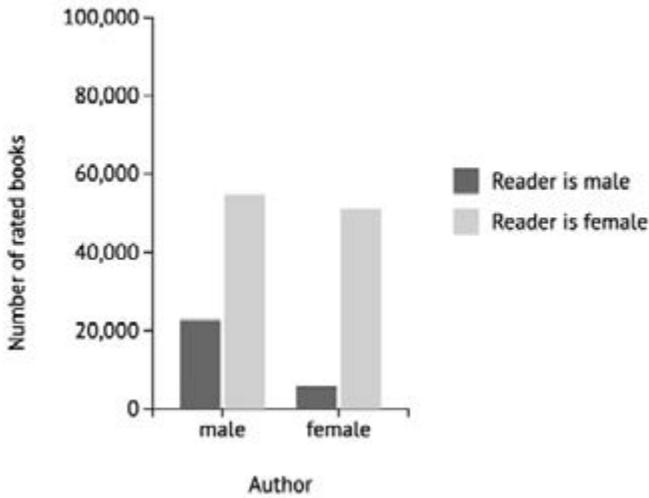


Figure 2.9: Number of books rated, by gender author and gender reader

Contrasting the motivations written about books belonging to different categories also yielded some interesting observations. Many participants referred to genre in their motivations, mostly when they were not very much impressed by a book's literary quality. Respondents described the rated book as, for example, a *women's novel*, a *doctor's novel*, a *holiday book*, or compared a book to a *Harlequin novel*. I contrasted the 3391 motivations about the 186 Suspense titles with those about the 215 books in the other three categories: Romance (with 773 motivations), Other (534 motivations), and Literary novel (7252 motivations). This added up to 8559 motivations. Negations were significantly more common in the motivations of Suspense: *no*, *not*, *few*. This suggested that this genre usually was not associated with high literary quality. Some nouns that are keywords in the motivations about Suspense referred to the genre (*thriller*, *detective*) where others refer to features that probably usually also were combined with a negation: *depth*, *character development*, *literature*, or to why a reader read the book: *relaxation*, *pastime*. The number of nouns characteristic of the other categories was considerably larger, but, probably because I combined the three other categories, they did not signal a unified trend. Some examples of keywords were *chick lit*, *composition*, *fantasy*, *events*, *history*, *intertextuality*, *children's book*, *life*, *originality*, *perspective*, *writing style*, *atmosphere*, *structure*, *symbolism*, *language*. Then there were a few author names that appeared more prominently. In the motivations of Suspense, they included the names of Dutch male author *Baantjer* and male translated authors *Baldacci*, *Brown*, *Grisham*, and *Mankell*, and only one female author (Dutch) *Loes den Hollander*. Keywords in the motivations of the other three categories were *Ammaniti*, *Irving*, *Japin*, *Murakami*, and

Spinoza – only men, including one from centuries ago. Spinoza is the protagonist of Irving D. Yalom's novel *Het raadsel Spinoza* (*The Spinoza Problem*) and it was therefore unsurprising to see readers of that novel mention that name in their motivations.

A final comparison concerns originals and translations. Of the 401 novels in the research corpus, 152 were originally written in Dutch. The other 249 had been translated into Dutch. We received a total of 4942 motivations for ratings of Dutch novels. The number of motivations for ratings of translated books was 7009. When contrasting these two sets of motivations, many personal names cropped up. These were mostly the names of the authors of the books listed, often used in a motivation for judging a book by that author, but sometimes other authors were mentioned. Qualifying adjectives used mainly for Dutch novels were *common, familiar, best, recognizable, beautiful, poetic, splendid, magnificent, careful*. More frequent in the motivations of translated novels were *deeper, feel good, long, long-winded, delicious, incredible, romantic, suspenseful, translated*. The country of origin of translations published in the Netherlands and the main source language were also mentioned more often: *American, English*. In the motivations about Dutch books the proportion of keywords that had to do with writing style and with close family ties was striking. On the other hand, the motivations about translated novels included more words referring to genres other than literary fiction. All in all, the results suggested that Dutch novels were judged more for their literary quality and translated novels more as genre fiction and for the suspense in the story. This may be due to the larger share of translated suspense and romantic-humorous books, but that should not be the only reason. After all, the corpus contains almost as many Dutch as translated novels in the category Literary novel.

Conclusion

This chapter described in detail how the Riddle team organized The National Reader Survey that ran in the Netherlands in 2013 and how the 401 books that participants were asked to share their opinion about were selected. In the response to the survey, we saw an overrepresentation of women, participants with vocational or university education, and participants of 65 years and older. Because these numbers align with what is known from other reading research, this overrepresentation is not problematic for research into the conventions of literariness. These were the actual people who read, bought, borrowed, and shared their opinions with other readers, professionally or not.

I also sketched the results of research using a representative panel of Dutch citizens that aimed to answer the question whether participants could be connected to one of two possible reader roles: a literary and 'distanced' one versus one mainly aiming for pleasure, described as 'identifying'. This resulted in the observation that almost all participants, whether they indicated to be very 'literary' readers or not, seemed to be

'identifying' readers, who for example wanted to be swept away by a novel. However, not all participants aspired a 'distanced' role, in which they for instance liked to explore the deeper layers of a novel. This more literary reader role therefore seemed to be an additional role a reader may or may not take on, but not an alternative role per se.

The chapter further presented the top 10 books most read and the top 10 most and least literary. The ten novels which received the highest scores for literary quality were all written by male authors and belonged to the category of Literary novels, and the ten receiving the lowest were all written by female authors, most from the Romance category. Because female authors also received significantly lower scores for literary quality for their Literary novels, I will analyse Literary novels written by female and by male authors each in separate chapters (4 and 5) in this book, only bringing them together in Chapter 6. As to the different categories or genres, the mean literary ratings showed a clear hierarchy: Literary novels ended up at the top, Romance at the bottom, and Suspense in between.

Many participants answered the question to motivate their score for literary quality in a free text field and I quoted elaborately from this rich source of information. I also compared different sets of motivations using keyword analysis to ascertain whether respondents used a different mix of words when describing books they rated with a high or a low score for literary quality, or whether the novel was a Dutch original or a translation from another language into Dutch. Furthermore, I compared the motivations written by female and male participants, older and younger respondents, and those with different levels of education. Upon analysis, the motivations yielded some general ideas about the reigning literary conventions in 2013, but not a lot of specifics yet. Most prominently mentioned were the writing style and deeper layers. Furthermore, literary quality seemed to be connected to 'difficult' or 'boring'. But which stylistic features exactly played a role in their judgement is hardly ever mentioned by the respondents. We will have to turn to the texts themselves to uncover this and this is what I will do in the next chapters of this book, starting with translations and some Romance and Suspense.

References

All hyperlinks were last accessed in February 2023.

Anthony, Laurence. *AntConc*. Version 3.5.8., Tokyo: Waseda University, 2019, <https://www.laurenceanthony.net/software.html>.

Culpeper, Jonathan. 'Keywords and Characterization. An Analysis of Six Characters in *Romeo and Juliet*.' David L. Hoover et al. *Digital Literary Studies. Corpus Approaches to Poetry, Prose, and Drama*. Routledge, 2014, pp. 9–34.

- Dixon, Peter, et al. 'Judging a Book by its Cover.' *Scientific Study of Literature*, vol. 5, no. 1, 2015, pp. 23–48, doi:10.1075/ssol.5.1.02dix.
- Felski, Rita. *Uses of Literature*. Oxford: Blackwell P, 2008.
- Koolen, Corina, et al. 'Literary Quality in the Eye of the Dutch Reader: The National Reader Survey.' *Poetics*, vol. 79, no. 101439, 2020, pp. 1–13, doi:10.1016/j.poetic.2020.101439.
- Riddell, Allen, and Karina van Dalen-Oskam, 'Readers and their Roles: Evidence from Readers of Contemporary Fiction in the Netherlands.' *PLoS ONE*, vol. 13, no. 7, 2018, pp. 1–12, doi:10.1371/journal.pone.0201157.
- Von Heydebrand, Renate, and Simone Winko. *Einführung in die Wertung von Literatur: Systematik – Geschichte – Legitimation* [Introduction to the Evaluation of Literature: Systematics – History – Legitimacy, not available in English]. Ferdinand Schöningh, 1996.

3. Romance, Suspense, and Translations

Mixed Methods in More Detail

Abstract: Translations were usually valued slightly less than Dutch originals. Participants attributed the lowest mean scores for literary quality to E.L. James' *Fifty Shades of Grey*. The Dutch publisher marketed the translations of the *Fifty Shades* trilogy as Literary novels. To find out if this label indeed suits the novels best, the chapter applies cluster analysis using Burrows' Delta and principal components analysis (PCA). The same methods are used to explore five novels by Swedish author Henning Mankell, to see whether the style of his Literary novels differs from his Suspense novels. Finally, keyword analysis is applied to get an indication of the fluency of translations, to see whether the results may explain the differences in ratings of originals and translations.

Keywords: Literary quality, *Fifty Shades* trilogy, Henning Mankell, Translation quality, Computational Literary Studies, Distant Reading

From Readers' Opinions to Text Analysis

This chapter presents the results of several experiments with the mixed-methods approach I selected with the aim to learn more about the conventions of literary quality in the Netherlands in 2013. It sketches the text analysis methods I applied, shows how they were combined with the results from The National Reader Survey described in the previous chapter, and outlines the ensuing observations. Most of the novels discussed in this chapter are translations and most belong to the categories of Romance and Suspense or, for one reason or another, skirt the lines between Literary novel, Suspense, and Romance. Special attention will be paid to the issue of translation quality.

In Chapter 2 I presented the top 10 most literary and least literary novels from the list of 401 books in the eyes of the Dutch readers who participated in the survey. As a reminder: The highest score a reader could attribute to a novel was 7, 'highly literary', the lowest 1, 'definitely not literary'. The novel to which readers attributed

the highest mean rating, Julian Barnes' *Alsof het voorbij is* (*The Sense of an Ending*) received an average of 6.6 and the one with the lowest, E.L. James' *Vijftig tinten grijs* (*Fifty Shades of Grey*) scored a 2.1. For all 401 books in the corpus the mean scores for both literary quality (the perceived literariness) and general quality (how good the book was considered to be) can be found in Appendix 2, ranked from highest score for literary quality to lowest. My first experiment was an exploratory analysis of these twenty novels applying a cluster analysis based on the hundred most frequent words. I selected the 'Classic Delta' measure that was based on Burrows' Delta (described in Chapter 1) and made use of the Stylo Package for R (Eder et al. 2016). In a cluster analysis, the most similar texts in terms of words and word frequencies are statistically linked in pairs after which the most similar clusters are also linked in pairs, and so on. The further to the right in the graph two books are connected, the more similar they are in the frequencies of the selected words – the graph is 'built' from right to left.

The result of this first measurement is visualized in Figure 3.1. The short names for each novel consist of the author name and, after the underscore, the first three words of the title that are not an article. The vertical line connects the books or clusters of books that are most similar in the frequencies of the hundred most used words. The author signal is clearly visible: all of the novels by Weisberger, by Mansell, and by Kinsella, all represented by more than one novel in the ten least literary novels on the list, cluster together in separate author clusters. Furthermore, the ten books with the lowest scores for literary quality (starting with L_, in a lighter colour in Figure 3.1) end up together in one of the two main clusters, and eight of the highest-scoring novels (starting with H_) are also clustered together. On the companion website to this book (see Appendix 3) this figure can be found in colour, added with visualizations from more measurements and with some more information.

For now, it suffices to note that for the selected hundred most frequent words, the novels *Norwegian Wood* (*Norwegian Wood*) by Haruki Murakami and *Alsof het voorbij is* (*The Sense of an Ending*) by Julian Barnes resembled the ten novels awarded the lowest literary quality slightly more than they resembled the other eight highly literary novels. An inspection of the hundred most frequent words in these twenty novels with the AntConc tool by Laurence Anthony showed only two of these are content words: the nouns 'eyes' and 'hand'. The other 98 are function words: articles, pronouns, prepositions, and verb forms of the verbs *to be* and of modal verbs. When I applied cluster analyses with more words (from 130 most frequent words onwards), Murakami and Barnes moved into the cluster with the highest literary quality, but they were joined by five of the least literary ones.

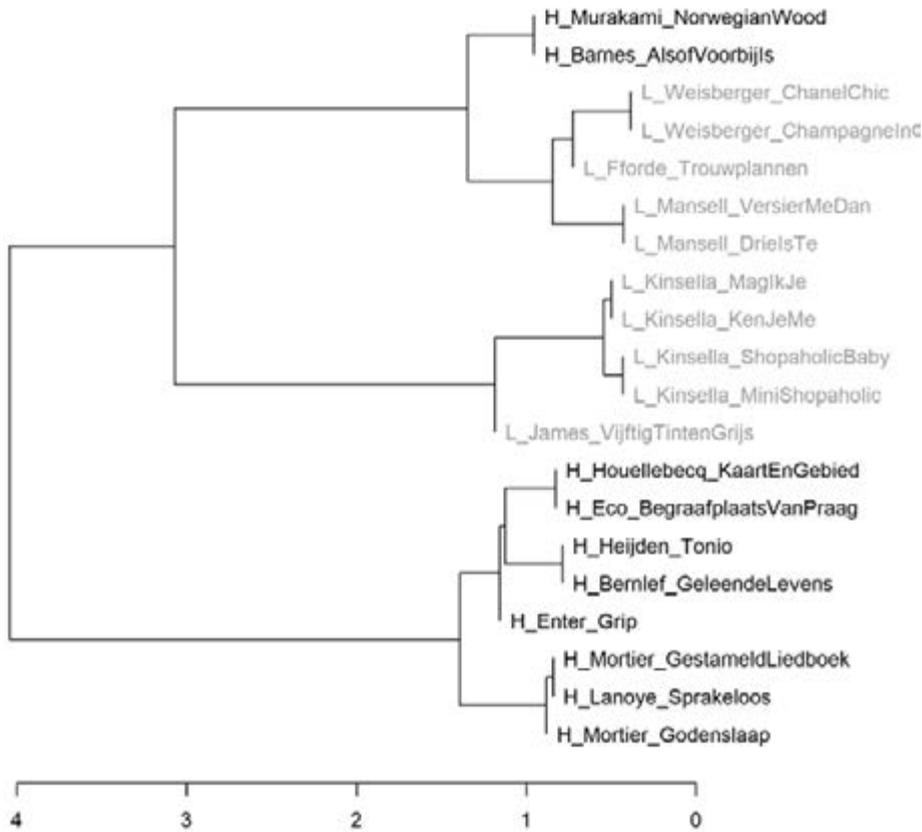


Figure 3.1: The ten novels with the highest (H) and lowest (L) scores for literary quality, Cluster Analysis (100 most frequent words)

Applying the same measurement to the fifty most literary and fifty least literary novels, which together took up a quarter of the complete corpus of 401 novels, yielded roughly the same results as for the ten most and least literary novels (graph not shown). More than forty books from both ends of the list clustered together. Barnes and Murakami here ended up in the cluster in which the highly literary books dominated. Here, seven of the novels with highest literary quality ended up in the cluster that included most of the novels that received the lowest scores for literary quality. They are listed below; an asterisk before the English titles indicates that the novel is not available in an English version. In those cases, I present an English translation of my own that is certainly not always the best possible translation, but will give at least some idea about the book to those who do not read Dutch.

These provisional translations are also included in the litRiddle package for R (see Appendix 3).

Anna Enquist, *De verdovers* (**The Anaesthetists*)
 Arnon Grunberg, *De man zonder ziekte* (**The Man Without Illness*)
 Arnon Grunberg, *Huid en haar* (**Hair and Hide*)
 P.F. Thomése, *De weldoener* (**The Benefactor*)
 Arthur Japin, *Dooi* (**Thaw*)
 Bernard Dewulf, *Kleine dagen* (**Little Days*)
 J.J. Voskuil, *De buurman* (**The Neighbour*)

The cluster dominated by novels that received the highest scores for literary quality included six books that belonged to the fifty novels with the lowest scores:

Jennifer Weiner, *Sommige meisjes* (*Certain Girls*)
 Candace Bushnell, *1 Fifth Avenue* (*One Fifth Avenue*)
 Mart Smeets, *Een koud kunstje* (**A Breeze*)
 Mart Smeets, *De afrekening* (**The Reckoning*)
 Suzanne Vermeer, *Cruise* (**Cruise*)
 Suzanne Vermeer, *Après-ski* (**Après Ski*)

Although there is again a measurable difference between most of the novels that received the highest and the lowest scores for literary quality, this difference does not explain yet how the novels differ exactly, nor why the readers rated them the way they did. An important point to note, however, is that the fifty most literary novels were almost all marketed as Literary novels, whereas most of the fifty least literary books had a different label. They included a few Suspense novels, five regional novels, and a lot of novels we categorized as Romance (more about the categorization in Chapter 2). It is therefore much more likely that the results of these first measurements did not mainly reflect the difference in literary quality but rather a difference in genre. They do not solve the riddle of literary quality but point the way to further explorations. Especially the outliers are intriguing and several of them will be discussed in more detail in the upcoming chapters.

Romance and the *Fifty Shades* Trilogy

The first books I will elaborate on are the three volumes of the *Fifty Shades* Trilogy. *Vijftig tinten grijs*, the translation of E.L. James' novel *Fifty Shades of Grey*, received

the lowest mean score for literary quality out of all 401 books. Participants who did not read this novel and did contribute an opinion about it, rated it even lower than those who did read it. Readers rated the literary quality of this first volume of the trilogy with an average of 2.1, the second and third volume both received a mean score of 2.6. The averages were determined based on a total of 627 ratings for the three volumes together of which the first one received the most, 280. Volume 2, *Vijftig tinten donkerder* (*Fifty Shades Darker*), received 169 ratings and Volume 3, *Vijftig tinten vrij* (*Fifty Shades Freed*), 179. Figure 3.2 shows that considerably more readers gave the very lowest score, 1, to the first volume of the trilogy than they gave to the second and third volume. Many of these readers probably did not continue reading volume 2 and 3. Those who did continue reading can be expected to have given higher ratings, with as a result the somewhat higher average scores for the second and third volume.

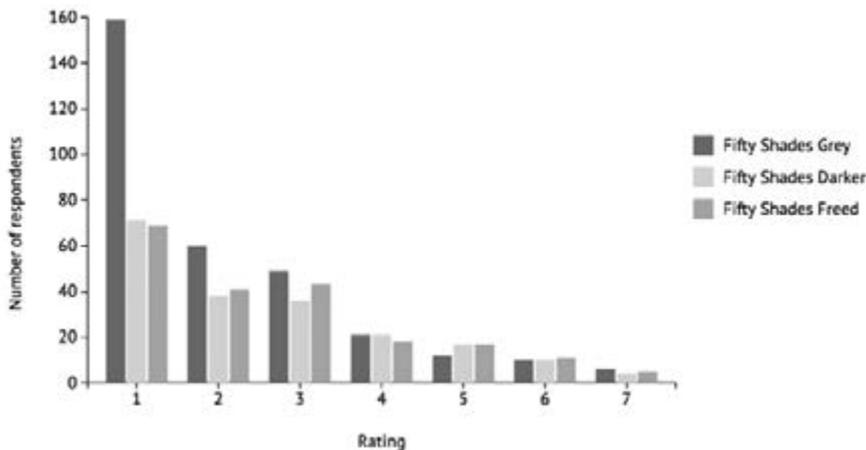


Figure 3.2: Ratings for the literary quality of E.L. James, *Vijftig tinten*-trilogie (*Fifty Shades Trilogy*)

To find out why readers were so disenchanted by the literary quality of the trilogy, I looked at the justifications readers provided for their rating. For the three novels together we recorded 143 motivations, nine of which were written by male participants. The words *niet* (not), *geen* (none), and *weinig* (little) occurred frequently but so did *veel* (many). Negative adjectives dominated: *slecht* (bad) and *simpel* (simple). The word *goed* (good) also frequently occurred, but as with *veel* (many), this may be immediately after a negation. Readers are referred to the litRiddle package for R if they would like to verify this themselves. I will quote a couple of the motivations as a teaser, starting with four of the very few respondents who gave one of the three volumes a score of 6 ('literary') or 7 ('highly literary'). I present my own English

translation of the motivations; the original Dutch text can be found in the litRiddle package for R by means of the added ID.

It is simply wonderful (Female, age 18; ID_975).

Because it is suspenseful, contains a story, and is written with love (Female, age 60; ID_12026).

Because it is well written since you are just completely drawn into the story (Female, age 16; ID_7676).

A story that really touches you (Female, 16; ID_5223).

The age of these respondents is striking: three very young participants (two aged sixteen and one aged eighteen – potential participants under sixteen were not allowed to take part in the survey) who indicated to be swept away by the novel and one aged sixty who appreciated the suspense of the story and remarked that it is ‘written with love’. Interesting as this may be, we cannot draw any solid conclusions from this precisely because the number of motivations for these high scores is so small. We may learn more from looking at motivations for the abundantly available low scores. These readers agreed about a poor writing style and a bad story. Some wondered aloud why they read it or why others read it.

The writing style is abominable, the story is devoid of substance, and it is obvious that the author has a very limited vocabulary. It is still a mystery to me why I read this book in the first place (Female, 36, score 2; ID_11795).

Language use, characters, and content. Lots of repetition of what happened, it was innovative because of the bondage aspect. More of an entertaining book and a hype (Female, 56, score 1; ID_5714).

Because not a single literary sentence can be found in the entire book! This book is only about the subject matter. Never read such a badly written book.

Content was a revelation but I don’t need to read a book for that (Female, 55, score 1; ID_2470).

As the second quote above shows a low score can go hand in hand with positive comments: the book was entertaining. Describing the novel as innovative was probably also meant positively. Something similar can be seen in the third quote, in which a reader admits that the content of the story was nothing less than a revelation to her – only to immediately undo the compliment.

Some readers found the emphasis on sex in the story sufficient reason to give it a low score for literary quality. Other readers wrote that the sex is well described:

Too much focus on sex (Female, 33, score 1; ID_13638).

Because the story seems closer to a book for older girls, rather than being original.

I do find the descriptions of sexual acts and physical things very good (Female, 62, score 3; ID_11347).

However, good descriptions of sex were not automatically perceived as literary descriptions:

This is absolute pulp, barely worth the paper it is written on. I don't understand why it is so highly praised. Fine what someone does in the bedroom but literature ... Marquis de Sade was literature, not this (Male, 54, score 1; ID_10479).

Another comparison was much more prominent in readers' substantiations of their opinions of the *Fifty Shades* trilogy. A great many readers compared the novels to popular romantic books on sale in supermarkets:

Ladies' novel. Even worse than the average Harlequin novel (Male, 53, score 1; ID_13048).

To me, this book feels like a more expensive edition of the Harlequin novels and other books about love and sex that can be found at the cash register in a supermarket (Female, 31, score 1; ID_11012).

In Chapter 2 I mentioned that the publisher of the Dutch translation of the three volumes of the *Fifty Shades* trilogy marketed the books with the label of Literary novel. Because the Riddle team was not convinced by this categorization, we decided to keep the novels apart from the other Literary novels in our corpus by putting them in our fourth category, Other. To find out whether the language of the trilogy is closer to the other Romance novels, as we suspected, or that they indeed are more alike to the other Literary novels in our corpus, I have taken several measurements, each time comparing the *Fifty Shades* trilogy with books from a different category. The results of this experiment are best illustrated by the visualizations of a principal components analysis (PCA) of the thousand most frequent words. PCA applied to the vocabulary of texts analyses correlations in the frequencies of the words used in the novels. It is a mathematically sophisticated procedure which maps the very complex abstract space of a large set of input variables (in this case: word frequencies) onto a conceptually simple, two-dimensional scatterplot. Such a visualization is much more intuitive for a researcher than multidimensional setups. It may be helpful to think of a designer trying to draw a building on a sheet of paper: the usual approach is to represent the most

important side (for example the front wall) as precisely as possible, while the side walls are deliberately transformed, to represent as much information as possible. After the transformation, the side wall coordinates are embedded into the X and Y dimensions. PCA performs a very similar transformation of the input variables, in order to identify the dominant information, namely PC1 and PC2. The other principal components can still be accessed and shown on a separate plot, but they usually contain very little information about the input data, and I refrain from showing them in the graphs in this book.

So, the first component, depicted on the horizontal axis in visualizations, reflects the cluster of correlated variables that together explain most of the variation in the data. Previous research on literary texts has found that the first principal component is usually related to the authorship of the texts. The second principal component, shown on the vertical axis, reflects the largest amount of correlating variables out of the remaining variables (i.e. those that do not correlate with the first component). Earlier research has shown that this can often be linked to a difference in genre or subgenre of the texts. More information about PCA in computational literary research can be found in Jockers (65–67 and 100–104), Hoover (26 and onwards) and Craig and Greatley-Hirsch (30–39). As before, I used the Stylo Package for R. For all PCAs I used the correlation version. In the visualizations of the measurements, books that are very similar are close together, and books that differ from the other selected novels end up as outliers. When exploring these visualizations it is important to remember that the results only show how the books that were selected for the experiment compare to each other. As soon as books are added or removed, the picture will change. Still, these visualizations may be of great help to get a rough idea about how the selected novels compare to each other in terms of most frequent words.

In figure 3.3, I have visualized the comparison of the trilogy with the 41 Romance titles. On the basis of a statistical comparison of vocabulary and word frequencies, the software has situated the trilogy far in its own corner compared to the other novels. Thus, the *Fifty Shades* trilogy is visibly different from the titles in the Romance category. It is important to remember that we did not have any representatives of the Harlequin type of novels in our corpus. To check whether the trilogy also deviates from this kind of popular romantic books we would need another corpus of texts, with representatives from the time period just before *Fifty Shades of Grey* took the book market by storm. For now, it is sufficient to note that the trilogy seems to be far removed from the novels in our corpus that have been labelled as Romance.

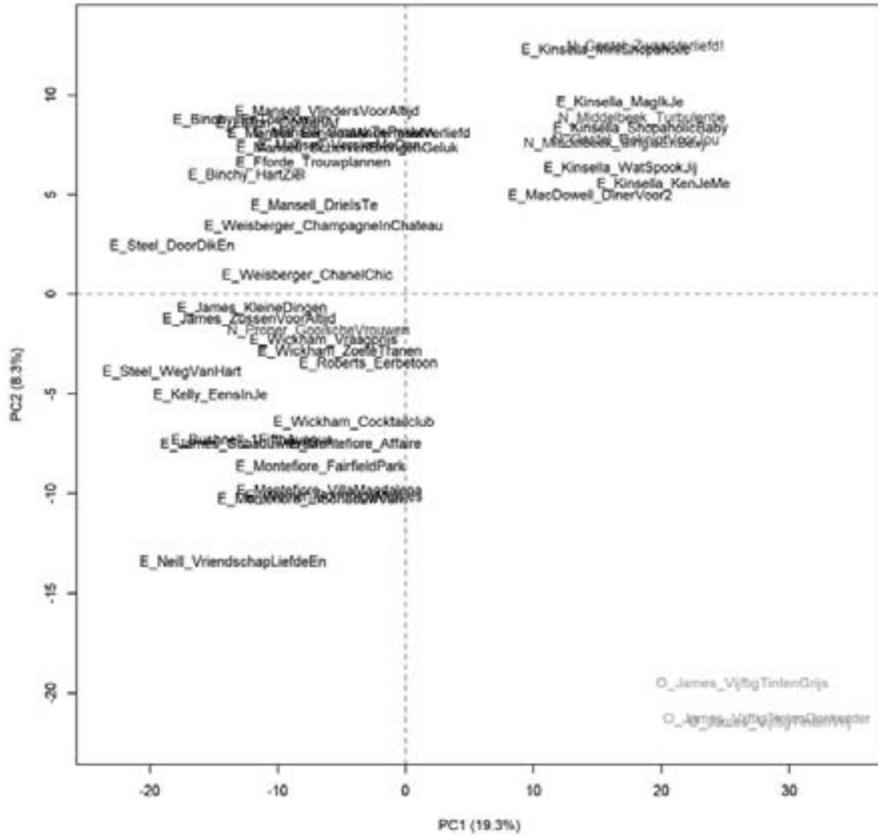


Figure 3.3: E.L. James, *Vijftig tinten*-trilogie (*Fifty Shades Trilogy*) and Romance, PCA (1000 most frequent words).

Novels translated from English are indicated with an E_ in front of the author name and shortened title and novels originally written in Dutch are indicated with N_. The O_ in front of the *Fifty Shades* titles stands for Other. The various groups have each been given their own shade of grey to make them easier to recognize.

Next, I compared the trilogy with Literary novels. For this, I did two sets of measurements, one comparing it with the Literary novels written by men and the other with those written by women – the visualizations would otherwise be too cluttered for an interpretable result. The graphs are not shown here, but they are available on the companion website to this book, see Appendix 3 for more information. Both times the trilogy stood well apart from the other novels. However, the distances are somewhat smaller than in the comparison with the Romance category. I would conclude that labelling the volumes of the trilogy as Literary novels is generous but perhaps not as far-fetched as we initially thought.

For the sake of completeness, I also included the third main genre, Suspense, in my experiments. Compared to Suspense written by men (see the visualization on the website) the trilogy again ended up clustering away from the rest of the titles. But the biggest surprise was the comparison of the trilogy with Suspense written by women, shown in Figure 3.4. There, the trilogy finally showed up close to other titles, even though only two. But those are two novels by Saskia Noort, known as the ‘queen of the Dutch thriller’.

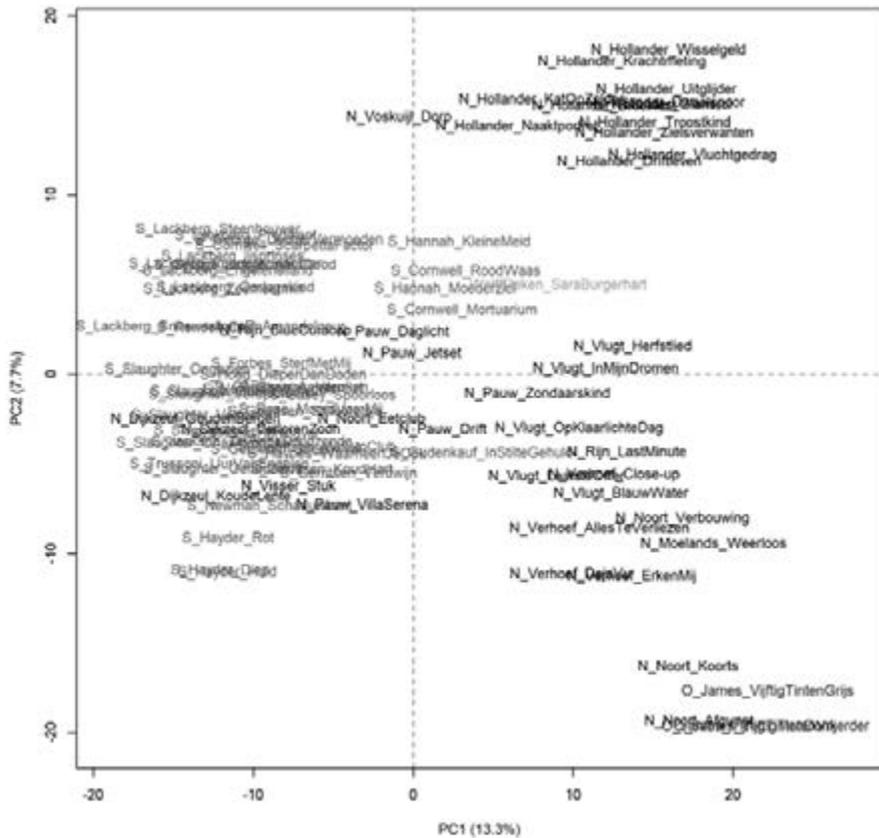


Figure 3.4: E.L. James, *Vijftig tinten*-trilogie (*Fifty Shades* Trilogy) and Suspense written by female authors, PCA (1000 most frequent words).

Novels translated from English are indicated with an S_ in front of the author name and shortened title and novels originally written in Dutch are indicated with N_. The O_ in front of the *Fifty Shades* titles stands for Other. The various groups have each been given their own shade of grey to make them easier to recognize.

Figure 3.4 is interesting for another reason: we can see a marked difference between the translations clustering on the left-hand side of the chart and most of

the originally Dutch thrillers on the right-hand side. Suspense by male authors did not show such a clear division, there Dutch novels and translations into Dutch from other languages happily mixed. The most probable explanation may be that a subgenre of Dutch thrillers written by female authors has developed in a very different direction compared to Dutch thrillers written by male authors and from translated thrillers, independent of the gender of the authors. It has become its own genre or subgenre. This observation is even more fascinating because the *Fifty Shades* trilogy joins the original Dutch thrillers and not the translations – to solidify this, more research is needed.

To contextualize this finding I will not go into scholarly discussions about Romance, feminism, and more – some helpful references in that direction may be found in Corina Koolen's PhD-thesis *Reading Beyond the Female: The Relationship Between Perception of Author Gender and Literary Quality* that will be discussed in more detail in Chapter 6. Instead, I turn to an essay by Dutch philosopher Marjan Slob titled *Foute fantasieën: Kleine filosofie van de ontvankelijkheid* (*Bad fantasies: Small Philosophy of Receptivity, not available in English). Not a scholarly publication indeed, but a personal, philosophical inquiry into why Slob enjoyed reading Romance novels – with an intriguing hypothesis that is very relevant for the *Fifty Shades* case and that would otherwise probably go unnoticed. I made use of the second edition (2017), where Slob in her introduction posited that the *Fifty Shades* trilogy seamlessly fits into the popular romantic genre. For her introduction to the genre Slob was indebted to the work of Janice A. Radway, *Reading the Romance: Women, Patriarchy, and Popular Literature* (1984/1991) and others. In her essay, Slob asked herself the following questions: What are these popular romantic books about? Why do women want to read them and men do not? Which fundamental difference between women and men, if any, can explain the difference in romantic reading choices? Slob stated:

The problem faced by the heroine in romantic fiction has still not been culturally processed: 'What does this attractive man want from me, what can he do with me, and how do I feel about that?' And socially, that question is as fundamental as ever for women worldwide. Yes, I really think that romantic fiction thematizes an issue of geopolitical proportions (14, my translation, cf. Van Dalen-Oskam [2021] 241).

Slob emphasized she knows that like women, men also ask themselves 'whether a woman wants more than a good time' (40). However, a 'pulp genre' around this question for men, with equally huge revenues worldwide as Harlequin has, does not exist.

I know that men are also vulnerable when they expose themselves to a woman. That they can feel cheated and used. That their heart can break. They too take

risks. But they will not think: how can I save myself if it turns out that the other person has bad intentions? If I am here naked, alone with him, and it goes awry? Because then you just won't win as a woman (107, cf. Van Dalen-Oskam [2021] 241).

According to Slob the underlying explanation for the difference in romantic reading choices is the non-trivial difference in physical strength – a secondary gender characteristic. Simply stated, men are physically stronger and this is a fact that women learn to take into account always and everywhere. Even in their fantasies. In fact, she assumed, in reading these romantic stories, women practise how to deal with this. According to Slob, the immense impact of this difference in physical strength explains why women and men have different fantasies. For (heterosexual) women, the fantasy apparently is the overwhelming man whom they can transform into the perfect partner.

The basic emotion of the romance would then be (erotic) rapture mixed with fear – despite all the nice and big words the hero and heroine mumble about love towards the end of the book. What is called a 'love story' is in reality a story of sublime shudders (120, cf. Van Dalen-Oskam [2021] 246).

Slob agreed with Radway and others that fantasies are not isolated. They arise from a particular view of reality and they are embedded in a continuous mutual influence between reality and fantasy. One's fantasies colour one's view of the world and one's view of the world in turn influences one's fantasies (26). Fantasies almost always lag behind the *desired* reality, so from a feminist perspective they can easily be labelled as 'bad' or politically incorrect. However, throughout time romantic stories change in ways that can be traced back to shifting attitudes and trends in society as a whole (62–63). For individuals, it may not be easy to just exchange bad fantasies for 'correct' ones. Slob emphasized that growing up in a certain culture usually means appropriating the dreams and fantasies of that culture. It is very difficult not to conform to what one's environment seems to expect of a person. When becoming a woman, one tries to accept what are generally supposed to be the desires of a 'real woman'. Slob characterized this process as 'autohypnosis' (158) and stated that the resulting fantasies can be changed only very minimally and in this way help bring about a change in the prevailing image. The changed image is again the source of dreams and fantasies of the next generation.

Girls feed on the images of existing women in their culture [...]. And boys feed on the images of men. In this way they come to desire what they are supposed to

desire. They come alive – that is to say, they go under cultural hypnosis. Culture reinforces the real differences between men and women. Perhaps that is not so strange; differences create tension, and eroticism thrives on that (163, cf. Van Dalen-Oskam [2021] 243).

In their motivations about the *Fifty Shades* trilogy participants in The National Reader Survey showed themselves to be readers who are familiar with the genre of popular romantic fiction but who have also learned, through ‘cultural hypnosis’, as Slob put it, to *not* appreciate that genre. They have learnt to deny it any literary quality. And even if they did enjoy reading the books, they were not willing to admit it. Perhaps this was not felt to be appropriate in the context of a large survey investigating the nature of literary quality, but it does say something about the then prevailing literary conventions.

The text analysis experiments I conducted showed that in its vocabulary, the *Fifty Shades* trilogy was most similar to a specific Dutch subgenre of Suspense written by women, even more than to the Literary novels in the corpus, and much more than to the books that have been labelled as Romance. As such, it aptly illustrates what the essay of philosopher Slob hypothesized, namely that suspense is one of the key ingredients of the genre of Harlequin novels – to which the *Fifty Shades* trilogy clearly belongs, according to Slob. Still, the readers’ ratings fit with the general ratings for the category of Romance and not with those for Suspense – or it would have received significantly higher scores.

The question remains why the publishing house opted for a counter-intuitive label for the trilogy – was the choice for Literary novels and not Romance (or Suspense?) deliberate, so as to make the books visible to readers who would otherwise not have noticed them in the bookstores – in other words: was it a marketing strategy in search of higher sales?

The solution to this riddle may lie in a concept from theoretical mathematics that is particularly relevant to marketing choices. Flemish researchers Jeroen Lauwers, Tom Deneire, and David Eelbode (Lauwers et al.) turned to game theory, a subfield of mathematics, to propose a new method for investigating which text properties could contribute to readers’ perception of literary quality. Their idea revolved around the Nash equilibrium, a concept that owes its name to mathematician John Forbes Nash, who described it in 1950. The starting point is a game with more than two participants, in which each participant independently determines the best strategy to achieve an optimal result. In economics the Nash equilibrium was applied, for example, in determining a strategy for entrepreneurs in their search for the best position in a given market. The surprising result was that the best position is exactly in the middle. That way, one can be sure to reach 50 per cent of all possible customers.

The three researchers previously applied it to rhetoric, the art of persuasion in spoken – but also in written – language (Deneire et al.). They argued that speakers who want to convince their audience must also look for a Nash equilibrium, so that here too, the best position is ‘in the middle’, between all possible opinions. Only when a speaker knows what the opinions of the target audience and possible other speakers are, can she decide which arguments she can best use in order to achieve her goal and convince her audience. The optimal strategy is to be as close to one’s opponent’s opinion as possible, but positioned slightly more in the middle. And if one doesn’t know what the opponent thinks, then the middle is the best position anyway.

The gap between rhetoric and politics on the one hand and literature on the other is smaller than it may seem. Lauwers, Deneire, and Eelbode assumed that the author of a literary work wants the reader to attribute literary quality to the written work. They argued that the literary quality of a text is based on an interplay of many different textual properties. To find out which of those properties outweigh others and what the ‘middle position’ of each property is, they proposed to question readers about their opinions on individual properties of a text and of the text as a whole. If this would show a correlation between the appreciation of the text as a whole and the preference for a certain feature among many readers, this might indicate that that particular feature is important to this audience in determining literary quality. Also, for that feature the middle position could be derived.

All in all, this method may seem to be somewhat similar to the research I report on in this book, but it is a lot more complex because readers would not only have to rate the novels as a whole, but would also be questioned about individual text features as such. It remains to be seen whether this approach is practically feasible on a scale that yields sufficient observations. Another concern is that Lauwers et al. focused on the perspective of the author and ignored the power of the publishers. I want to argue that it is precisely here that the Nash equilibrium can play a role. Publishers do their utmost to position the novels they publish in the literary field in such a way as to attract as many of their target customers as possible. This influence lies, among other things, in the attribution of a genre label. The computer analysis of the vocabulary and word frequencies showed that the *Fifty Shades* trilogy is far removed from those titles in our research corpus that are labelled Romance by the publishers, and strikingly similar to some novels labelled Suspense written by one of the female Dutch authors. The trilogy did come near the Literary novel, but still remained at a certain distance. Yet, given the enormous revenues the books must have generated and the overwhelming attention to them, that very label seems to have been the ideal ‘middle ground’. With this label, the publisher may have found the balance between minimal risk and maximum profit.

Crossing the Borders of Suspense

The genre fluidity of the *Fifty Shades* trilogy not only involved the Literary novel and Romance but, perhaps surprisingly, also Suspense. Almost half of the corpus of 401 novels, 186 titles, were labelled Suspense. The number of books authored by female and male writers in this category did not radically differ and the same can be said for the mean scores for literary quality and general quality the participants of The National Reader Survey attributed to them. One could therefore say that, pertaining to gender of authors and ratings of books written by female and male authors, this genre is the most equal. The books labelled Romance were dominated by female authors and female readers and received the lowest scores for literary quality, whereas the category of Literary novels was dominated by male authors and was being read by a mixed group of readers, who awarded it the highest ratings. I will not extensively deal with the Suspense genre and only focus on one very interesting case where it operates on the borderline with the Literary novel.

Table 3.1: Five novels by Henning Mankell

Genre	Literary quality	General quality	Title Dutch translation / year of publication	Translator	Title Swedish, year of publication [English translation]
Literary novel	5.00	5.78	De Daisy sisters (2009)	Edith Sybesma	Daisy Sisters (1982) [not available in English]
Literary novel	4.92	5.97	De geschiedenis van een gevallen engel (2011)	Clementine Luijten	Minnnet av en smutsig ängel (2011) [A Treacherous Paradise]
Suspense	4.83	5.89	Kennedy's brein (2007)	Carole Post van der Linde	Kennedys hjäma (2005) [Kennedy's Brain]
Suspense	4.75	5.97	De Chinees (2009)	Corry van Bree	Kinesen (2008) [The Man from Beijing]
Suspense	4.77	6.00	De gekwelde man (2010)	Janny Middelbeek-Oortgiesen	Den orolige mannen (2009) [The Troubled Man]

The Swedish author Henning Mankell (1948–2015) is perhaps best known for his suspenseful books, but he is also known as a literary author. He was the creator of policeman Kurt Wallander, who is the protagonist of thirteen of his seventeen Suspense novels. He has also written children's books and fourteen literary novels. Three of his thrillers and two of his literary novels have ended up in our corpus (Table 3.1). *De Daisy Sisters* (**The Daisy Sisters*) was first published in Dutch in 2009, translated by Edith Sybesma from the Swedish *Daisy Sisters*, which in turn was originally published in 1982. The novel narrates the difficult lives of Eivor Maria Skoglund, her

mother, and her daughter. The other literary novel by Henning Mankell in our corpus is *Minnet av en smutsig ängel*, which appeared in Swedish in 2011 and was translated into Dutch by Clementine Luijten the same year as *De geschiedenis van een gevallen engel* (*A Treacherous Paradise*). Here, too, a woman is the protagonist of the story. The novel tells how Hanna Renström in 1904 moves into a hotel in Lourenço Marquez, which turns out to be a brothel, and how she copes with this environment. The three Suspense novels by Mankell that met the criteria for inclusion in our corpus (most sold from bookstores and/or most borrowed from public libraries in 2009–2012 and first published in Dutch in 2007 or after) are *Kennedy's brein* (*Kennedy's Brain*), *De Chinees* (*The Man from Beijing*), and, from the Kurt Wallander series, *De gekwelde man* (*The Troubled Man*). *Kennedy's brein* and *De Chinees* also have female protagonists.

The five novels by Mankell have each been translated into Dutch by a different translator (see Table 3.1). All five books were rated by between 440 to 504 readers. The average scores for literary quality were between 4.75 and 5.00. The two literary novels received the highest scores but they were not that much higher than the scores for the three Suspense novels. These three ended remarkably high up in the literary ranking, wedged in between many other books from the category of the Literary novel. This leads to the question whether Mankell's Suspense novels are perceived to be more like Literary novels than other books categorized as Suspense.

In a small experiment, I analysed all forty translated Literary novels by male authors in our corpus, which include Mankell's. To this selection I added his three Suspense novels. A whole series of PCA analyses yielded almost the same outcome. As an example, I show the visualization of the result of the PCA of the thousand most frequent words (Figure 3.5). Mankell's three Suspense novels appear somewhat separately from the other novels, together with his Literary novel *De geschiedenis van een gevallen engel*. However, his other Literary novel *De Daisy Sisters* is very far removed from the other 4.

When I added more thrillers to the comparison, this pattern did not change. The Suspense novels by, for example, Stieg Larsson and Jussi Adler-Olsen then took up the space between the Mankell cluster at the bottom of the graph and the rest of the field, being more similar to the other literary novels than the four books by Mankell (see the website for visualizations of these additional measurements). The conclusion is that, in terms of vocabulary and word frequency, *De geschiedenis van een gevallen engel* (*A Treacherous Paradise*) is close to the Suspense novels that Mankell wrote in the same time period. It seems, therefore, that although Mankell published books in two different genres, Suspense and Literary novel, his style as reflected in his use of words, perhaps is not very different. The fact that *De Daisy Sisters* was an outlier does not disprove this because the original is much older (1982) and it is known that an author's writing style may develop over time just like languages and the conventions that apply to different genres, as David Hoover (2021) recently demonstrated. Further research into Mankell's complete oeuvre would

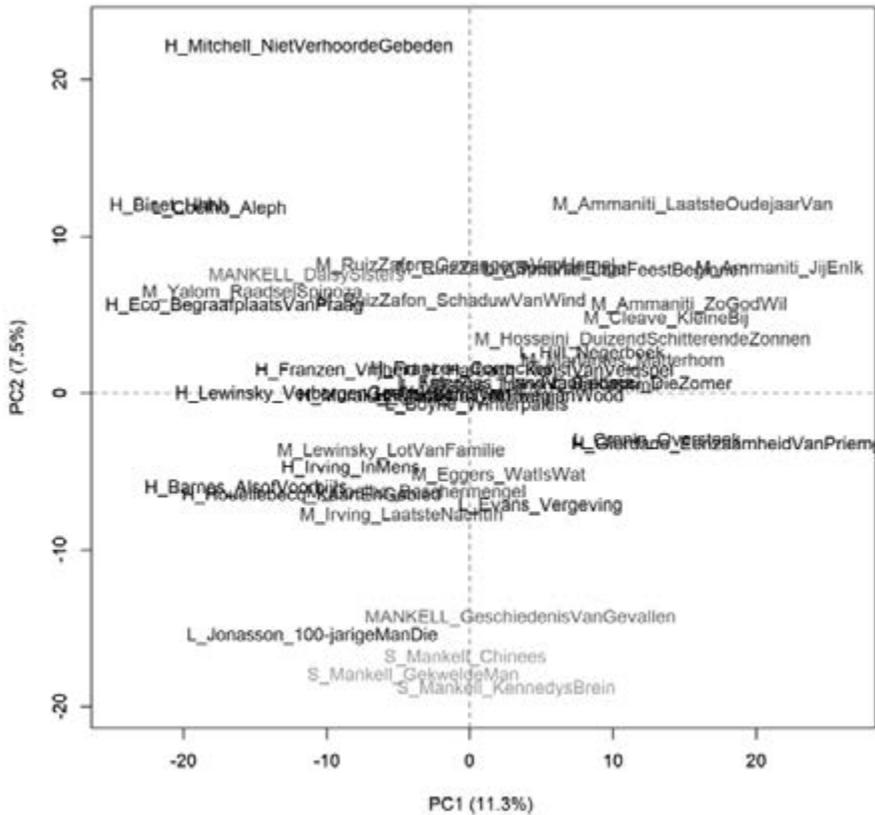


Figure 3.5: Translated Literary novels written by male authors and three Suspense novels by Mankell, PCA (1000 most frequent words).

Mankell's Suspense novels are indicated with an S_ in front of the author name and shortened title. The Literary novels are divided in those with the highest scores for Literary quality (H_), the lowest scores (L_) and the intermediate group (M_). The various groups have each been given their own shade of grey to make them easier to recognize.

be needed to confirm this. If this discrepancy between genre label and Mankell's writing style is correct, what is interesting is that such a development can even be found in translations. This probably reflects current translation conventions in the Dutch language area – and also needs further research.

One of the Literary novels that shifted position after I added Mankell's Suspense to the mix is *De 100-jarige man die uit het raam klom en verdween*, the Dutch translation of the debut novel of Swedish author Jonas Jonasson, *Hundraåringen som klev ut genom fönstret och försvann* (*The 100-Year-Old Man Who Climbed Out the Window and Disappeared*). The Dutch version was released in 2011 and was a smash hit. It was translated by Corry van Bree, who was also responsible for the translation of Mankell's

De Chinees (The Man from Beijing). The story is just as unusual as its title. The 99-year-old Allan Karlsson is bored but not bored to death, to his great annoyance. When his hundredth birthday arrives and the administrators of his elderly home have prepared a party for him, he quietly climbs out of his window and sets off to the nearest bus station in his slippers. A slapstick world trip full of car chases and explosions ensues. Perhaps these explosions explain why, in terms of vocabulary and word frequency, the novel most resembles the thrillers among which it ended up in my last measurements.

To return to Mankell, it is not clear how his reputation influenced readers in their ratings of the literary quality of these five books. Did his Suspense novels get higher scores than those of other Suspense authors because readers knew Mankell also wrote Literary novels? They did end up between the Literary novels in the ranked list of all 401 books. Or would the mean scores for the two Literary novels have even been higher if Mankell would not have been well-known as a Suspense author? These questions cannot be answered for now.

The Riddle of Translation Quality

The Mankell case also leads us to another question: Are the ratings the participants in The National Reader Survey attributed to translated novels similar to novels originally written in Dutch? Many Dutch readers also read other languages than Dutch, as can be seen time and again in their answer to the question ‘Can you explain why you rate this book as you do?’

I read the book in **Spanish**, beautiful diction! (ID_3137, about Carlos Ruiz Zafón, *Het spel van de engel [The Angel's Game]*).

Read it in **French**. Wonderful use of language and topical issues very originally and subtly presented. (ID_5763, about Michel Houellebecq, *De kaart en het gebied [The Map and the Territory]*).

Read it in **Danish**, master it (Danish) sufficiently to find the book suspenseful but cannot give a literary judgement (ID_4967, about Jussi Adler-Olsen, *De bedrijfsterrorist [*The Corporate Terrorist]*).

The story is very important in this book. The language a little less so. Nevertheless, the books in the Potter series are full of wordplay and hidden references that are largely lost in Dutch. NB: That's why I read this book in **English!** (ID_3985, about J.K. Rowling, *Harry Potter en de Relieken van de Dood [Harry Potter and the Deathly Hallows]*).

As I already indicated in Chapter 1 of this monograph, translations play an important role in the Netherlands. Researcher Thomas Franssen calculated that of all poetry

and fiction titles published in 2009, 62.8 per cent were translations. Most translations came from English and belonged to the category of literary novels, many more were translated from different languages than in the categories of suspense and romance (Franssen 30–37). The Riddle research corpus shows the same distribution. Of the 249 translated books in our corpus of 401, 180 were translated from English, out of which 104 originated in the United States and 67 in the United Kingdom. The remaining 69 came – in order of most to least number of titles – from Swedish, Italian, Spanish, Danish, Norwegian, French, German, Icelandic, Japanese, Portuguese, and Afrikaans. I suspect that many respondents based their assessment of a book on reading it in its original language, but I do not think this problematic. Their opinions were still welcome.

The previous chapter showed that participants in The National Reader Survey regularly referred to the fact that a book was a translation when substantiating their rating for literary quality. Older participants (sixty years and over) did so slightly more often than younger ones (sixteen–forty-five years). Many respondents were well aware that they were reading a translation. They regularly praised the translated book, but sometimes they were less enthusiastic or suspended a negative judgement about the literary quality by suggesting that the problem may also lie with the translation:

Read the **translation**, which was unreadable because of the grammatically incorrect sentences, and therefore did not finish the book. Maybe I would have had a different opinion if I had read the original English edition (ID_5434, about Kathryn Stockett, *Een keukenmeidenroman* [*The Help*]).

Very beautifully written, good plot. Good **translation** (ID_7312, about Tatiana de Rosnay, *Die laatste zomer* [*A Secret Kept*]).

Maybe it was the bad **translation**, but the writing style was not exactly literary (ID_5465, about Silvia Avallone, *Staal* [*Swimming to Elba*]).

Apart from a suspenseful plot the book has deeper layers for the attentive reader. Moreover, the author made extensive use of intertextuality and other so-called literary procedures. And not unimportant: much care was taken with the Dutch **translation**; it strikes me that this is not always the case with books I consider less literary (e.g. thrillers, which I often detest for that reason) (ID_10902, about Carlos Ruiz Zafón, *De schaduw van de wind* [*The Shadow of the Wind*]).

Story a bit over the top. Not always beautiful language. (Can also be due to the **translation** of course!) (ID_11472, about John Boyne, *Het winterpaleis* [*The House Of Special Purpose*]).

On the seven-point scale from 1 to 7, the 152 novels written in Dutch received an average score of 4.95 for literary quality, just above the average of 4.67 for all 401 novels, while the 249 translated novels ended slightly below that average, with a

score of 4.47. The lower score for translations may largely be due to the fact that the categories of Suspense and Romance contain considerably more translations than original Dutch novels and both are genres that are awarded less literary ratings than titles from the category Literary novel anyway.

Translations did not receive similar scores for literary quality as novels written in Dutch in the same category. The scores for Dutch and translated Suspense were not that far apart, while the average score for translated Literary novels is lower than that for original Dutch Literary novels (Table 3.3). That category also was the only one with an almost equal number of titles, namely 73 Dutch and 74 translated Literary novels (Table 3.2). In the genres of Suspense and Romance, on the other hand, the translated books collectively received a higher score than the Dutch ones. The differences are statistically significant; they can be easily verified with the litRiddle package for R (see Appendix 3).

Table 3.2: Categories and number of books originally in Dutch and translated into Dutch

Category	Dutch	Translated into Dutch	Total
Literary novel	73	74	147
Suspense	58	128	186
Romance	5	36	41
Other	16	11	27

Table 3.3: Categories in Dutch and translated with mean score for literary quality

Category	Literary quality	Category	Literary quality
Dutch	4.95	Translated	4.47
Literary novel Dutch	5.52	Literary novel translated	5.35
Suspense Dutch	3.76	Suspense translated	4.02
Romance Dutch	2.75	Romance translated	2.97
Other Dutch	4.74	Other translated	3.54

Table 3.4: Categories of Literary novel and Suspense in Dutch and translated with mean score for literary quality

Novels	Number of Literary novels	Literary quality	Number of Suspense novels	Literary quality
Dutch	73	5.52	58	3.76
Translated from English	44	5.16	89	3.91
Translated from other languages than English	30	5.56	39	4.19

Table 3.4 shows that in the categories of Literary novel and Suspense, novels translated from English on average scored slightly lower than books translated from other languages. Translated Suspense from languages other than English even scored significantly higher than those written in Dutch. But in the Dutch literary field, as in our corpus, we found far fewer translations from languages other than English. It may be that it is precisely this relative rarity that makes translations from languages other than English seem slightly more valuable in the eyes of readers.

Another explanation should also be considered, however: it may in fact be the case that translations from English are simply worse than those from other languages. And it would be extremely useful if we could somehow easily measure whether readers have a point in their criticism of a translation. Earlier in this chapter I quoted a reader stating a translation had ‘grammatically incorrect sentences’. Another participant in The National Reader Survey wrote: ‘Bad book, too American, unrealistic, badly translated, ugly sentences’ (ID_8793). These readers did not explain what exactly is incorrect or ugly about these sentences but that does not make their opinions any less relevant.

Translation scholars have given considerable thought to what makes a good translation. Below, I cite only a few of their publications in English. First and foremost, they agree that the answer to this question is not easy. After all, translation is not simply replacing a word in one language (the source language) with an equivalent in another language (the target language). Languages and language systems are never exactly the same and a translation can therefore never be completely identical to the original. What a translator necessarily does is interpreting the source text and reformulating it in the target language, also taking into account the cultural context (Bassnett 3). Moreover, the ideas of what constitutes a good translation change over time and this underlines there is no fixed standard for a successful translation (Bassnett 81). Little research has been done into measurable aspects of translation (cf. Lewandowska-Tomaszczyk 6).

Negative comments about translations are mainly made by people who know the original language well enough to think they can see that the translation is inadequate, according to translation scholar David Bellos in his book *Is That a Fish in your Ear? The Amazing Adventure of Translation* (36). This statement seems to fit perfectly with the audience that took part in The National Reader Survey, given the motivations cited earlier in this chapter in which readers explicitly stated that they had read a book in the original language. The most widely read and spoken language in the Netherlands other than Dutch is English and on the list of 401 novels, the majority of translations came from English. French, German, and many other languages followed at a great distance. If we are to believe Bellos, then the widespread knowledge of English may contribute to the observation that Dutch

readers seemed to trust translations from English less than translations from other languages. As I showed in the previous section, the average scores for literary quality indeed give the impression that it is mainly translations from English that readers find fault with (see Table 3.4). I therefore examined these in more detail in two experiments. I made use of the knowledge that many readers consider a translation to be good if the language use does not reveal that it is a translation, or if the translation reads as if it were an original. The term that translators and translation scholars use for this is ‘fluency’; a text can be experienced as a fluent translation from the perspective of the receiving language and culture. In my experiments I limit my focus to linguistic features, assuming that the degree of fluency of a translation could be revealed by comparing the vocabulary of translated novels with original Dutch novels.

The subject of the first experiment is a set of Dutch words that linguists call ‘modal particles’. These are common words that carry little meaning in themselves but which can be used, for example, to diminish or cushion an announcement. In a Dutch sentence such as ‘Ga maar even zitten’ (‘Just sit down’) the words *maar* and *even* can be described as modal particles. They are difficult to translate into English. In addition to *maar* and *even* we may also include *dan*, *eens*, *misschien*, *nou*, *soms*, and *toch*. English does not know them and has other ways to achieve the same goal. And that has consequences for Dutch-English and English-Dutch translators: they must consciously choose other constructions to express the same intention. Translation scholars have noted that texts translated from English into Dutch contain fewer particles than texts written in Dutch. Together with linguist Ton van der Wouden and data scientist Saskia Lensink I investigated whether this is also the case in the books translated from English in the research corpus of 401 novels (Lensink et al.). We focused on the categories Literary novel, Suspense, and Romance, and examined the occurrence of the modal particles in 305 novels, 136 written in Dutch and 169 translated from English. The measurements showed that the modal particles *dan*, *ook*, *even*, *toch*, and *eens* occurred significantly more often in novels written in Dutch. The modal particles *maar*, *nou*, *soms*, and *misschien* appeared significantly more often in novels translated from English into Dutch.

For the Dutch modal particles *even*, *toch*, and *eens*, we also found a difference in usage between the various genres. Using an algorithm based on the occurrence of these words, we wanted to predict whether a novel was a translation or written in Dutch. It turned out that it was not possible to recognize translated books in the category of Literary novel, but that it was possible to recognize translations in the categories of Suspense and Romance. This indicates that in the use of these three modal particles, the translations from English into Dutch in the category Literary novel can be called fluent – the frequency of use was comparable to that

in the original Dutch literary novels. This also meant that the translations from the genres of Suspense and Romance were significantly less fluent in the use of the modal particles *even*, *toch*, and *eens* than those from the category of the Literary novel. I consider it quite possible that readers subconsciously noticed this and got the feeling that a translation was substandard.

In the second experiment I applied a method I have used before in this book: comparing the word frequencies of all words in two text collections in order to find out which words are significantly more frequent in one group compared to the other. I used the keyword function of the AntConc tool developed by Laurence Anthony. For this purpose, I used a smaller selection of novels translated from English, namely from the category Literary novels written by women – 23 Dutch and 25 novels translated from English. Comparing the frequencies of the thousand most frequent words in these two groups with the help of text analysis software did not show a clear dichotomy. However, a comparison of the vocabulary in both groups did yield words that occurred significantly more often than might be expected in one of the two groups. On average, the 23 Dutch novels contained considerably more often the personal pronouns *ik* (I), *me/mij* (me), and *mijn* (mine). The translations contained *ze* (she), *haar* (her), and *hen* (them) relatively more often. This seems to have little to do with translation quality. It may be that the selected Dutch novels more often presented a first-person narration while the novels translated from English were written more often in a third-person perspective. This was confirmed by research conducted by Lisanne van Rossum in her master's thesis at Utrecht University in 2020.

In the 25 novels translated from English, a number of verb forms occurred significantly more often in the translated than in the original Dutch novels: *vroeg* (asked), *wist* (knew), and *zei* (said). These are often used to mark speech in dialogue. The pattern may also be coincidental, for instance because the English originals of the selected books contained more dialogue than the Dutch novels. Another possibility is that it was not the proportion of dialogue that differed, but the way in which it was introduced, for instance when it would be more usual in original Dutch novels of the same genre to omit such an indication. In that case, we could conclude that here we have encountered a reflection of different literary conventions in the two languages. Yet another possible explanation is that translators may have added the dialogue marks themselves from time to time. The English originals, like many Dutch novels, then did not always explicitly signal the interactions in a dialogue with a verb. Research has shown that translators often tend to render the text they translate more understandable and accessible, among other things by adding these kinds of words. For an example of changing dialogue markers in Swedish translations from English, see Bellos (188). If translators indeed added such dialogue markers,

I would argue that the translation can be labelled as less fluent. The Dutch translation may be seamless, but the text has been made easier to read than was probably intended in the original. This simplification may not fit the genre to which the novel belongs.

The 25 translations also used the conjunctions *en* (and), *toen* (then), *dat* (that), *voordat* (before), *terwijl* (while), and *hoewel* (although) significantly more often. I suspect that the more frequent use of *terwijl* (while) in the translations has to do with the English participle constructions (such as *being*, *walking*, etc.), which in Dutch is very unusual and in translation is usually replaced by a subordinate clause, beginning with, for instance, *terwijl*. And *hoewel* (although) is probably the translation of the fairly frequent *though* or *although* in English. The word *hoewel* is particularly striking in this list, because for Dutch readers the word is archaic.¹ I would conclude this is a good example of word usage that is less fluent from a Dutch perspective. For the more frequent conjunctions *en* (and), *toen* (then), *dat* (that), and *voordat* (before) in the translations, this conclusion is more difficult to draw. As with the more frequently used *vroeg* (asked), *wist* (knew), and *zei* (said), translators may have unconsciously added them to make the story easier to follow. This can be seen as a characteristic of a less fluent translation. Books in the Suspense category also had significantly more conjunctions in the translations from English than in the original Dutch books, which confirms that genre differences may play a role. For now, we can conclude that if we start with the idea that a good translation features fluent Dutch and is not recognizable as such, it seems that in some respects (the modal particles), translated literary novels in our corpus are slightly more fluent than novels labelled as Suspense or Romance. In other respects, however, genres show no differences (the conjunctions, especially *terwijl* [while] and *hoewel* [although]).

To verify these observations I turned to another statistical approach that I used earlier in this chapter: principal components analysis (PCA) using the Stylo package for R. I use the visualizations of the results of the calculations for the first two components as an aid to see how the selected novels compare to each other. The largest difference is depicted on the horizontal axis, the first component, and the second largest on the vertical axis. Figure 3.6 shows a visualization of a PCA of the thousand most frequent words. Titles printed in black (starting with E_) are translations from English and titles printed in grey (starting with N_) are written in Dutch. In this figure the Dutch novels and novels translated from English into Dutch are scattered across the graph, largely overlapping. This means that they cannot be clearly distinguished from each other in their use of

¹ *Hoewel* is one of the words on the list of archaic words of the Genootschap Onze Taal, the Dutch language society, see <https://onzetaal.nl/taalloket/modern-taalgebruik>.

the thousand most frequent words. The picture changes, however, if we make a PCA of only the six conjunctions that in the keyword approach sketched above I found to be especially frequent in translations from English. Figure 3.7 illustrates that novels written in Dutch and novels translated from English hardly overlap and occupy two clearly distinct areas in the graph. This confirms that there is a significant difference in the use of these six words in the originals and the translations.

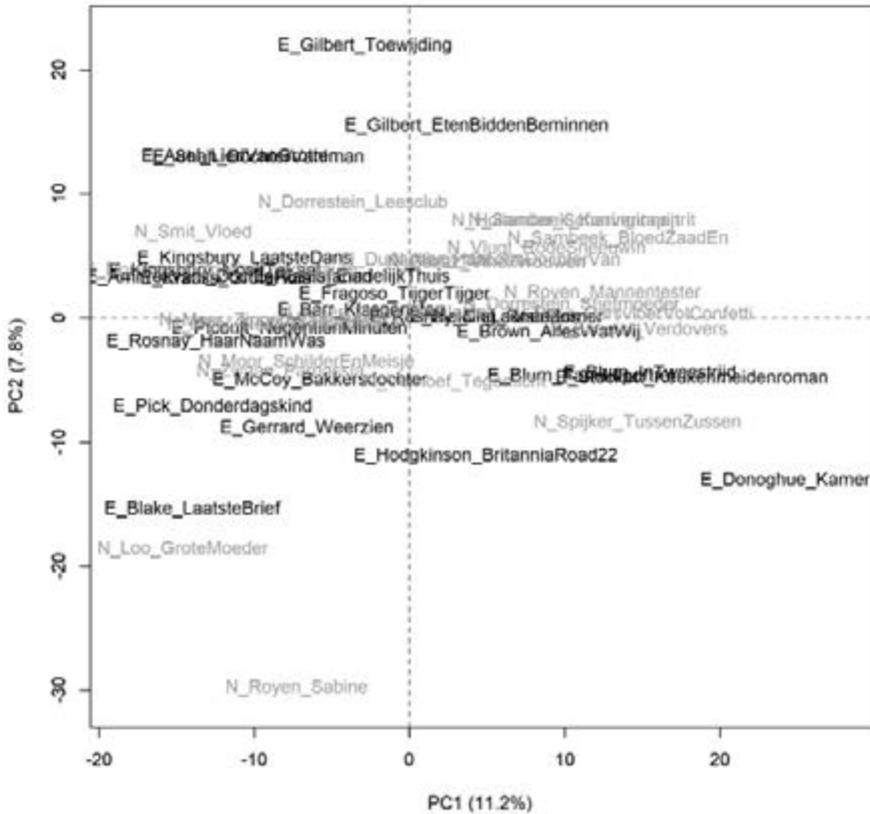


Figure 3.6: Literary novels by female authors written in Dutch and translated from English, PCA (1000 most frequent words).
 Novels translated from English are indicated with an E_ in front of the author name and shortened title and novels originally written in Dutch are indicated with N_. The various groups have each been given their own shade of grey to make them easier to recognize.

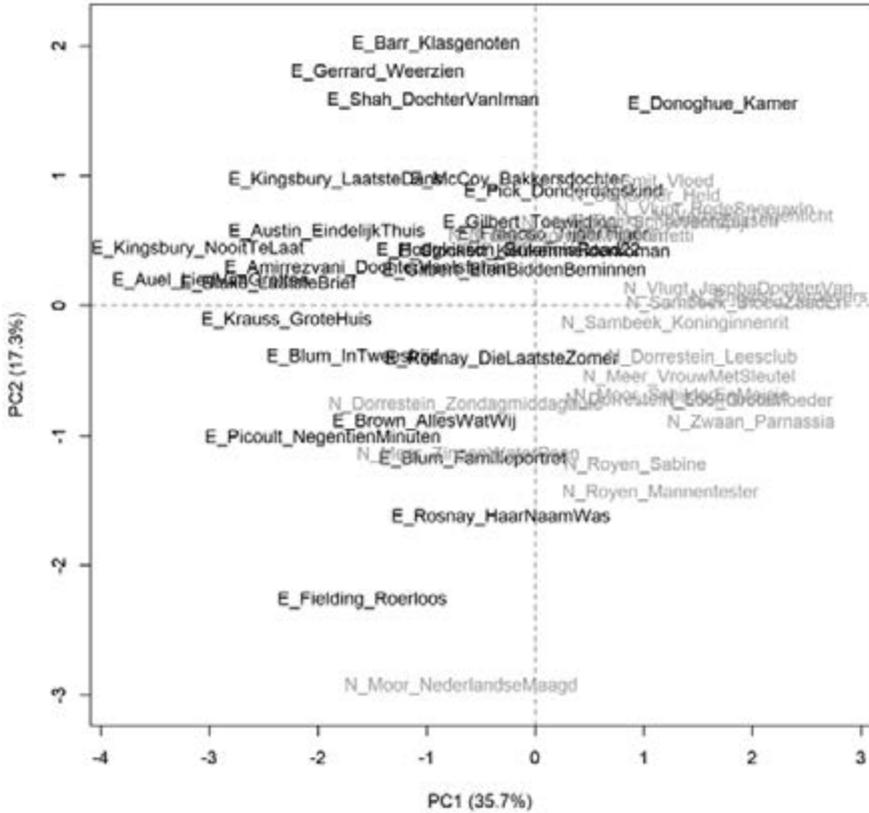


Figure 3.7: The conjunctions *en* (and), *toen* (then), *dat* (that), *voordat* (before), *terwijl* (while), and *hoewel* (although) in Literary novels by female authors written in Dutch and translated from English, PCA. Novels translated from English are indicated with an E_ in front of the author name and shortened title and novels originally written in Dutch are indicated with N_. The two groups have each been given their own shade of grey to make them easier to recognize.

Two novels in Figure 3.7 seem to be outliers compared to the rest of their group. *Kamer* (*Room*) by Emma Donoghue at the top right of the graph is the exception among the translated novels that we find scattered across the left half. The Dutch translation of Donoghue's novel is closer to the Literary novels written in Dutch. Checking the occurrences of all six words in the novel itself, I found that the conjunction *hoewel* (although) is not used even once. I will return to this novel in the next chapter. Among the Dutch novels *De Nederlandse maagd* (*The Dutch Maiden*) at the bottom left of the graph claims a place in the margin. Further investigation revealed that one of the reasons for this may be the relatively frequent use of the conjunction *hoewel* (although) compared to the other Dutch novels. The conjunction

hoewel was found 10611 times in the 401 novels in the research corpus. That is far too many to peruse with the naked eye. In the 23 Dutch Literary novels and the 25 translations of English Literary novels written by female authors it is found 1221 times. The number of occurrences in the Dutch novels is 179 and in the translations from English 1042 – almost six times as many as in the Dutch novels. Using the concordance function in AntConc I viewed all occurrences in their respective contexts, sorting the first word after *hoewel* alphabetically. This made it fairly easy to determine whether there were major differences in usage between the originally Dutch novels and the translations from English. It seemed to me that the way the word is used is indeed largely similar, only with many more occurrences in the translations. The exception among the Dutch novels, Marente de Moor's *De Nederlandse maagd* (*The Dutch Maiden*), has sixteen of the 179 occurrences in all 23 Dutch Literary novels written by women. As I remarked earlier, Dutch *hoewel* is archaic for speakers and readers of Dutch. It is very well possible that Marente de Moor used it consciously or unconsciously as one of the stylistic devices to emphasize that her novel is set in the past.

Based on the experiments described above I am convinced that we can indeed measure some characteristics of fluency in translation. Here, I have dealt with only some of them. Many more linguistic features may play a role. For now, the properties I have discussed here can be used as a kind of yardstick: they give an indication of the possible degree of fluency of different translations. And my preliminary conclusion is that they only partially vindicate the participants in The National Reader Survey who complained about translation quality.

Translations in the Periphery

Participants in The National Reader Survey seemed to be very much aware of reading a translation and this awareness may have influenced their expectations of the translation quality of a novel. If they read a Suspense novel, they might have expected that the translation would not be very high-quality. Any criticism they had on the translation quality had little impact on the scores they gave to Suspense novels, which were comparable to the scores they gave to Dutch Suspense. When they read a Literary novel in a Dutch translation, we did see a lower score on literary quality for Literary novels translated from English, than for novels written in Dutch.

I also attempted some very basic ways to measure the fluency of translations. Roughly put, a fluent translation is experienced by the reader as if it were originally written in Dutch. I have shown that in the corpus of 401 novels many translations from English did not fully meet this description. However, there was a clear genre difference. Suspense translated from English had a measurably lower fluency than

Literary novels translated from English. Dutch translations of English-language Literary novels also showed deviations compared to novels written in Dutch, but less so than those in the category of Suspense. However, Dutch readers rated the literary quality of those translations somewhat lower than that of the original Dutch literary novels. I therefore suspect that for readers and non-readers the fact that a novel *is* a translation – no matter how good – already detracts from the expected and perceived literary quality.

Translations of Literary novels thus seem to be judged based on characteristics that do not apply for literary quality in Suspense. This emphasizes once again that readers expect attention to language, style, word usage, and sentence structure to be a core element of literary novels. If ‘literary’ language is not appropriately rendered in Dutch, then part of what makes the novel literary is lost. Language and translation status are therefore indeed part of the conventions that make a novel more or less literary in the eyes of the readers.

On top of all this, prestige probably played a role. We saw that Literary novels translated from languages other than English received a higher score for literary quality than those translated from English. Probably without being aware of it, the participants in The National Reader Survey judged translations of Literary novels more harshly than novels written in Dutch, and even were stricter for translations from English than those from other languages. Translation scholars such as Bassnett and Venuti emphasized that currently, the Anglo-American culture that is the origin of so many translations in our research corpus, is the dominant cultural power worldwide. Researchers have explained this situation with the so-called centre-periphery model that I introduced in the first chapter of this book (Bevens et al.) and which I will briefly review here. Bevers et al. located the centre of the contemporary international art world in the United States and several Western-European countries. Around that centre, a few countries and regions occupy a semi-central or semi-peripheral position, and the rest of the world is in the periphery. In the global literary field, the Netherlands plays a role in the periphery (13, 380–381). Here, what is in the centre generally receives quite a lot of attention, whereas in the centre, the periphery is hardly visible. In this context, Bevers and colleagues also referred to the Matthew Effect: wealthy people tend to acquire more wealth, whereas poor people get poorer. They also pointed to what is known as ‘the law of small numbers’ to explain why usually such a small number of books is extremely successful while so many other books are largely ignored and not successful at all. Because only a limited number of authors can be in the spotlight at any one time, publishers, critics, and readers keep falling back on the same names with the result that the number of successful writers remains relatively small. For writers, Bevers et al. identified a mechanism of ‘self-reinforcing reputation’: ‘Once achieved, success opens doors, provides credit, more publicity, which in turn strengthens the reputation’ (Bevens et al. 16–17, 386).

In 2013, novels in English or translated from English were the most present and visible on the Dutch market. They received by far the most attention from publishers and therefore also from buyers, borrowers, and readers. Approached from the perspective of the law of small numbers, there was little room left for novels in other languages. It was probably much harder for those authors and books to obtain the same attention. But perhaps when they were selected for publication and appeared in the bookstores, they stood out exactly *because* of the prevalence of English. Their relative rarity may have given these books more prestige, which would then explain the higher literary appreciation among the participants in The National Reader Survey. Another factor that may have played a role in the relatively higher literary quality ascribed to literary translations from languages other than English is that, if something is in high supply, its quality will also be considerably more diverse. This applies both to translations from English and to the proficiency of the people who produce them. There were many more translations from English and many more translators from English than translators from other languages. It is not surprising then should we encounter more inadequate translations from English than from other languages.

In my opinion there is something else at play here as well. Dutch secondary schools also teach students one or multiple other modern languages than Dutch, and that fact alone will have confronted them with language differences and translation issues. Translation scholar Bassnett pointed out that precisely in times and cultures where the complexity of translation is recognized, confidence in translation seems to decline. As soon as a translation is equated with a certain loss in meaning, the translation acquires a lower status than the original (Bassnett 11). I would argue that in most schools, this process may have been actively supported: students of a foreign language were taught to distrust translations by definition. This might have influenced how students feel about translations later as adult readers, with the regrettable side effect that the work of translators remained underappreciated.

Conclusion

This chapter reported on the results of several text analysis experiments. A comparison of the most and least literary novels from the Riddle corpus showed that most of the novels clustered together with others that received a comparably high or low rating for literary quality. This did not solve the riddle of literary quality, however, because the influence of genre conventions was probably much larger than that of conventions of literariness. To uncover the linguistic differences between highly valued and lowly valued novels it is therefore necessary to exclude genre as a confounding factor from these experiments by focusing solely on the

category of the Literary novel. Because of the significant differences in literary value attributed to novels written by female and by male authors, I will keep these separate and analyse each in more detail in their own chapter. In Chapter 4 I will first go into Literary novels written by female authors, and in Chapter 5 into those written by male authors. In Chapter 6, I will reintroduce them for a final experiment, next to a further discussion of the gender element based on Corina Koolen's work and an overview of the results of high-level measurements done by Andreas van Cranenburgh.

Genre also played a role in two other experiments. In the first I tried to find out which of the main categories (Literary novel, Suspense, or Romance) the *Fifty Shades* trilogy showed most similarities to in terms of vocabulary and word frequencies. The trilogy, labelled by the Dutch publisher as Literary novels, proved to be least close to the novels labelled Romance, and closest to Suspense written by one of the Dutch authors in the corpus. The second experiment addressed the five titles by Swedish author Henning Mankell, two of which were Literary novels and three were Suspense novels. One of the Literary novels firmly clustered with the three Suspense novels, but the other one ended up very far away from the other four. This probably can be explained by the fact that this was a translation of a novel published much earlier in Swedish and may therefore reflect an earlier phase in Mankell's development as a writer.

The other main topic of this chapter was whether differences in translation quality may have informed differences in ratings of the literary quality of translations. I introduced two methods to measure elements of the level of fluency of a translation, which showed that generally, Literary novels resulted in a somewhat more fluent translation than Suspense. Furthermore, readers did unconsciously seem to take into account in their ratings whether a novel was translated or not, but the actual quality of the translation was less important for the ratings of Suspense novels than for Literary novels. And in the last case, readers trusted translations from English less than those from other languages, from which translations were much rarer. Because readers may have been slightly biased against translations, I will keep them separated from Dutch originals in the next chapters.

References

All hyperlinks were last accessed in February 2023.

Anthony, Laurence. *AntConc*. Version 3.5.8., Tokyo: Waseda University, 2019, www.laurenceanthony.net/software.

Bassnett, Susan. *Translation*. Routledge, 2014.

- Bellos, David. *Is That a Fish in your Ear? The Amazing Adventure of Translation*. 2011. Penguin Books, 2012.
- Bevers, Ton, et al. *Nederlandse kunst in de wereld: Literatuur, architectuur en beeldende kunst 1980-2013* [*Dutch Art in the World: Literature, Architecture and Visual Arts 1980-2013*, not available in English]. Vantilt, 2015.
- Calvo Tello, José. *The Novel in the Spanish Silver Age. A Digital Analysis of Genre Using Machine Learning*. Bielefeld University Press, 2021, doi:10.14361/9783839459256.
- Craig, Hugh, and Brett Greatley-Hirsch. *Style, Computers, and Early Modern Drama: Beyond Authorship*. Cambridge University Press, 2017.
- Deneire, Tom, et al. 'A Game with Words: Rhetorical Citizenship and Game Theory'. *Contemporary Rhetorical Citizenship*, edited by Christian Kock and Lisa Villadsen. Leiden University Press, 2014, pp. 323–340.
- Eder, Maciej, et al. 'Stylometry with R: A Package for Computational Text Analysis'. *The R Journal*, vol. 8, no. 1, 2016, pp. 107–121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.
- Franssen, Thomas. *How Books Travel: Translation Flows and Practices of Dutch Acquiring Editors and New York Literary Scouts, 1980-2009*. 2014. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245/1.446961>.
- Hoover, David L. *Modes of Composition and the Durability of Style in Literature*. Routledge, 2021.
- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013.
- Koolen, Corina. *Reading Beyond the Female: The Relationship Between Perception of Author Gender and Literary Quality*. 2018. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245/1/cb936704-8215-4f47-9013-0d43d37f1ce7>.
- Lauwers, Jeroen, et al. 'The Question of Literary Quality: A Rhetorical and Game-Theoretical Approach'. *Scientific Study of Literature*, vol. 5, no. 1, 2015, pp. 3–22, doi: 10.1075/ssol.5.1.01lau.
- Lensink, Saskia, et al. 'Hidden Indicators: Modal Particles in Modern Dutch Fiction and Fiction Translated from English, A Quantitative Approach'. Unpublished manuscript, https://tonvanderwouden.nl/index_files/papers/Lensink-Wouden-Dalen.pdf.
- Lewandowska-Tomaszczyk, Barbara. 'Explicit and Tacit: An Interplay of the Quantitative and Qualitative Approaches to Translation'. *Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research*, edited by Michael P. Oakes en Meng Ji. John Benjamins Publishing Company, 2012, pp. 3–34.
- Oakes, Michael P., and Meng Ji, editors. *Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research*, Studies in Corpus Linguistics 51. John Benjamins Publishing Company, 2012.

- Radway, Janice A. *Reading the Romance: Women, Patriarchy, and Popular Literature*. 1984. Updated ed., University of North Carolina Press, 1991.
- Slob, Marjan. *Foute fantasieën: Kleine filosofie van de ontvankelijkheid*. [*Bad Fantasies: Small Philosophy of Receptivity*, not available in English]. 2007. Updated ed., Rotterdam: Lemniscaat, 2017.
- Van Cranenburgh, Andreas. *Rich Statistical Parsing and Literary Language*. 2016. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37f1ce7>.
- Van Dalen-Oskam, Karina. *Het raadsel literatuur: Is literaire kwaliteit meetbaar? [The Riddle of Literature: Can We Measure Literary Quality?]* Amsterdam University Press, 2021, <https://karinavdo.github.io/RaadselLiteratuur/>.
- Van Rossum, Lisanne M. *Towards Literary Narrative Perspective: The Relationship Between Narrative Perspective and Perceived Literariness*. 2020. Utrecht University, research master's thesis. *Utrecht University Repository*, <https://studenttheses.uu.nl/handle/20.500.12932/38257>.
- Venuti, Lawrence. *The Translator's Invisibility: A History of Translation*. 1995. 2nd ed., Routledge, 2008.

4. Literary Novels Written by Women

Donoghue and Dorrestein

Abstract: Literary novels written by female authors received lower scores for literary quality than those written by male authors. In this chapter, the literary novels written by women are analysed, starting with translations and then followed by Dutch originals. Cluster analysis using Burrows' Delta, principal components analysis (PCA), and keyword analysis are applied to explore differences in word usage and word frequencies between highly rated and lowly rated novels in order to find out whether these may explain the ratings. Among the translations, Emma Donoghue's *Room* and Elizabeth Gilbert's *Eat, Pray, Love* receive special attention. Works by Dutch authors Margriet de Moor, Marente de Moor, Naima El-Bezaz, and Renate Dorrestein are described in more detail.

Keywords: Literary quality, Emma Donoghue, Elizabeth Gilbert, Renate Dorrestein, Computational Literary Studies, Distant Reading

The Glass Ceiling

The previous chapter showed that literary quality in the eyes of the Dutch readers who participated in The National Reader Survey in 2013 was strongly linked to the perceived genre of a novel. Books labelled Romance in the research corpus received the lowest scores, those marketed as Literary novels generally received the highest scores, and Suspense sat in between. Continuing my quest in search of the conventions of literariness in the Netherlands in 2013, I will now focus on the category of Literary novels. As became clear in Chapter 3, Literary novels by female authors generally received lower scores than those by male authors, and the fact that a novel was a translation may also have influenced readers' ratings, sometimes negatively and sometimes positively. For these reasons, I will describe novels written by female and male authors in separate chapters. The first of these two chapters is dedicated to the Literary novels by female authors. I have included

a novel co-authored by two women but excluded a novel by a mixed duo. After a brief glance at the position of female authors in the literary ranking, I will first zoom in on the 34 translations and then on the 23 Literary novels written in Dutch by women.

The research corpus of 401 novels contained a total of 196 books with a female author. Of these, we labelled 57 as Literary novels, 83 as Suspense, 40 as Romance and 16 as Other. With all 401 novels ranked according to their mean score for literary quality, the 'glass ceiling' that still exists in Dutch society can be found between floor 31 and 32. The first 31 ranks are taken by Literary novels written by male authors, and the 32nd place has the highest ranking Literary novel written by a woman: *Het grote huis* (2010) by Nicole Krauss, translated from the (American) English *Great House* (2010) by Tjadine Stheeman and Rob van der Veer. Literary novels by female authors are mainly found between rank 32 and rank 136 – in Appendix 2 the complete list of books can be found, ordered from highest literary quality to lowest. Suspense written by women predominantly occurs between ranks 140 and 304. Romance, mainly written by female authors except for one title written by a mixed duo, ended up between ranks 331 and 400. Literary novels also occasionally occur lower down on the ranked list, and some titles from the Suspense and Romance categories can be found higher up than in the ranges I mentioned above. It is striking that the Dutch Suspense novel *Dorp* (**Village*; the asterisk indicates no English translation is available and that the English title is my own provisional translation) by Anouschka Voskuil is on rank 65, right in the middle of the Literary novels, and that we find the romantic *Diner voor 2* (*Turning Tables*) by female duo Heather and Rose MacDowell on place 133, close to the Literary novels and still above most of the Suspense. *Diner voor 2* was rated by 41 participants, which is somewhat below the minimum of 50 ratings that we would have liked to have for each book. *Dorp* was rated by 48 respondents, also just below the minimum. This slightly lower number of ratings may or may not have played a part in the exceptional position in the ranking, and therefore I will not be tempted to draw any conclusions about their ranking.

A general glance at the list of the Literary novels written by female authors shows that the upper half contains more Dutch novels, and the lower half more translations from English. Translations from other languages can be found in different parts of the ranked list. Three novels translated from Spanish ended up in the top half, together with one translation from German and one from Italian. A second translation from German is on the negative side of the ranking. Three translations from French feature in the middle range, together with a set of translations from English. It is also striking that the four novels considered least literary are Dutch. Those who would like to explore the list in more detail are referred to Appendix 2 and to the litRiddle package for R (see Appendix 3).

Nicole Krauss' *Het grote huis* (*Great House*) narrates several life stories, in all of which the same antique writing desk plays a role. In The National Reader Survey, 483 readers rated the novel. The average of their scores for literary quality was 5.99 on a scale of 7. Most of them gave it a 6 (literary) for literary quality, followed by sevens (highly literary) and fives (somewhat literary) – the readers were unanimous in their praise (see Figure 4.1).

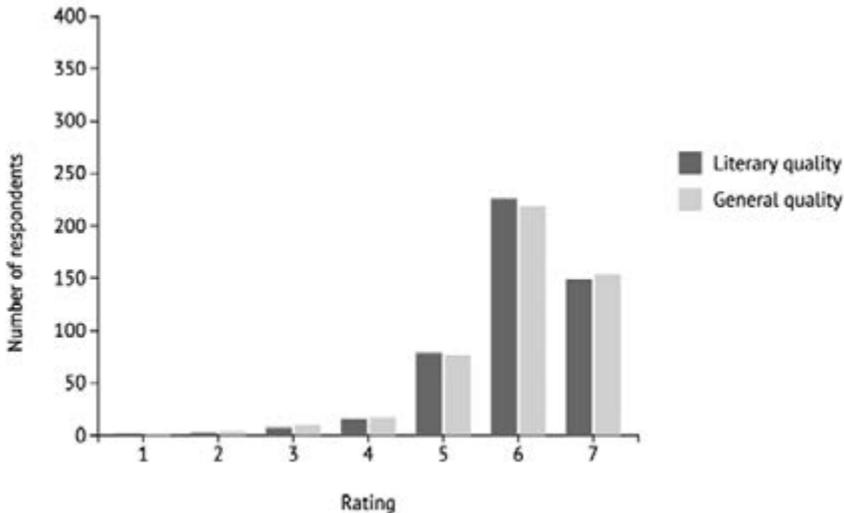


Figure 4.1: Ratings of Nicole Krauss, *Het grote huis* (*Great House*)

A total of 76 respondents presented us with some more information. Many of them referred to the language and the deeper layers in the novel to substantiate their ratings. What is striking about these motivations is that several participants reported that it was a book that begged to be reread. To quote one of them – those who would like to read the original Dutch can find it in the litRiddle R package through the ID number presented here by way of reference: ‘Wonderful writing style, beautiful story lines. A book to reread several times, then you get more and more out of it’ (ID_6381). Another reader wrote: ‘Read it in my book club and we discussed it in depth’ (ID_5366), perhaps suggesting that a book that lends itself to an in-depth discussion by definition has high literary quality. A final striking motivation I would like to quote is the following: ‘She clearly adopted her husband’s writing style, but to me literature is where good writing and a good story come together’ (ID_4077). The first part of this statement written by a female respondent seems rather accusatory – Krauss supposedly ‘nicked’ the writing style of her (ex) husband, Jonathan Safran Foer – author of, among other books, *Everything is Illuminated* and *Extremely Loud and Incredibly Close*. But the result is very good.

The respondent is positive, but not wholeheartedly so. Foer is not on our list, by the way.

Another Literary novel translated from English ended up remarkably far on the negative side of the ranking. Shortly after its publication, the book was widely discussed and debated, but this apparently did not help it gain literary prestige. *Eten, bidden, beminnen* by Elizabeth Gilbert received an average score of 3.51 for literary quality. The full title of Elizabeth Gilbert's book is *Eten, bidden, beminnen. De zoektocht van een vrouw in Italië, India en Indonesië*. It appeared in 2007 in the Dutch translation by Martine Jellema after the English *Eat, Pray, Love: One Woman's Search for Everything across Italy, India and Indonesia* that was published in 2006. The publishing house of the Dutch translation labelled it as a translated literary novel. The book is presented as the author's personal account of a year in which she tries to get her life back into balance. Just over six hundred participants in The National Reader Survey who indicated that they had read the book rated it on both literary and general quality, with widely varying scores, as Figure 4.2 shows. When it comes to general quality – how good do you think this book is? – the scores selected most often were 5 and 6. For literary quality, 5 was also the most frequently given score, but the most frequent after that was a rating of 3.

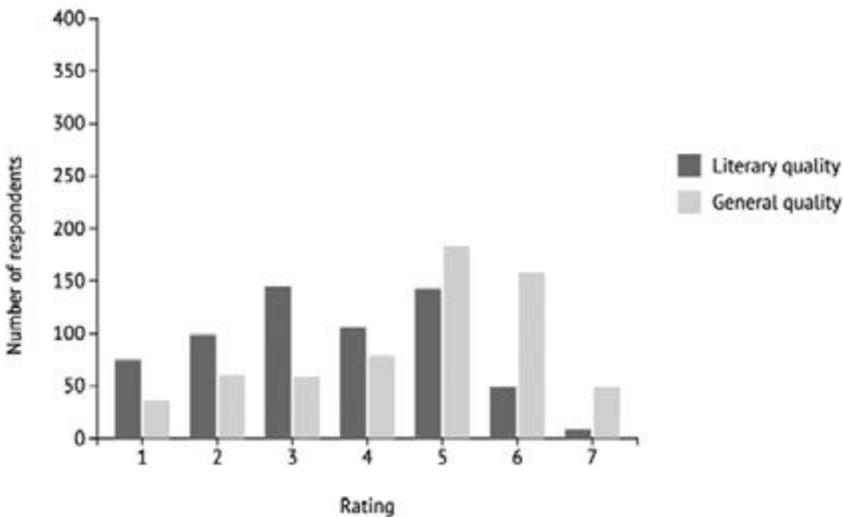


Figure 4.2: Ratings of Elizabeth Gilbert, *Eten, bidden, beminnen* (*Eat, Pray, Love*)

A total of 89 readers motivated their rating, all but four of whom were women. Below I present a small selection; for a list of all substantiations I again refer to the litRiddle package for R, suggesting to additionally use a translation tool to

get a general idea about their nature in your own preferred language. The ratings are indicated before the quote, from score (1), definitely not literary, to (7), highly literary. The highest score was represented only once with a motivation; most of the substantiations were for low ratings for literary quality.

- (1) It simply is a true story of a woman in a personal crisis (ID_213).
- (1) The style was cringe-inducing (ID_8840).
- (1) What a boring, contrived, predictable book. I'm quite willing to empathize with other people, other perspectives on life, and other experiences, but this was so packed with platitudes that after doing my best for forty pages I diagonally read the rest at random. So I didn't finish reading the book, which is very much against my habit (no matter how bad a book is) and it went straight to the second-hand bookstore in almost new condition! (ID_9474).
- (2) This book is nothing more than the lament of a woman in her midlife crisis. What a wimp that Elizabeth is (ID_7596).
- (2) Thought it rather resembled a Harlequin novel (ID_6310).
- (2) I liked the book a lot, it made me very happy, but it is written in a rather 'straightforward' way. I read it because of the interesting story, but not for deep insights or beautiful language (ID_4875).
- (3) Nicely entertaining book, not delving very deep (ID_4794).
- (3) The book has a straightforward style, it does not have any deeper layer. The language is not very good (ID_12986).
- (4) The story is predictable. It is not well written. Typical American whining story (ID_3802).
- (4) Actually, it is not a literary novel. It is an autobiographical work packaged as a novel. A cheap novel, which could not captivate me (ID_1289).
- (5) The main character grows as a person throughout the book. She actually understands less at the beginning than at the end (ID_9596).
- (6) Different from other novels. Especially because different aspects of cultures and religions are described (ID_13582).
- (7) The book is written in neat sentences and is at times profound (ID_12312).

The motivations for their opinions about the literary quality seem to revolve around a genre aspect. Readers sometimes described Gilbert's novel as non-fiction, as autobiographical, and therefore as not being literary. Repeatedly it is characterized as a women's book, a ladies' novel, a kitchen maid's novel, or a Harlequin novel. Furthermore, the novel was mostly experienced as not profound, not beautifully written, and as not having any deeper layers. It was easy to read, chatty, a simple story. It was too popular to be literature, too American, and too superficial. And the narrator, Elizabeth, was having a midlife crisis and called simply a 'wimp'. Is the distance between the

opinions about Krauss' highly rated novel and those about Gilbert's loathed book reflected in the texts themselves, analysed by means of text analysis tools?

Translations Compared

To find out how the 34 translated Literary novels written by female authors compare to each other in their vocabulary use I divided them into three groups: those with the highest mean scores for literary quality, those with the lowest scores, and the group in between.

Group H: highest scores for literary quality – mean score

- Nicole Krauss, *Het grote huis* (*Great House*) – 5.99
- Julia Franck, *De middagvrouw* (*The Blind Side of the Heart*) – 5.73
- Isabel Allende, *Het eiland onder de zee* (*Island Beneath the Sea*) – 5.64
- Isabel Allende, *Het negende schrift van Maya* (*Maya's Notebook*) – 5.50
- Emma Donoghue, *Kamer* (*Room*) – 5.45
- Silvia Avallone, *Staal* (*Swimming to Elba*) – 5.45
- María Dueñas, *Het geluid van de nacht* (*The Time in Between*) – 5.37
- Alison Pick, *Donderdagskind* (*Far to Go*) – 4.97
- Kathryn Stockett, *Een keukenmeidenroman* (*The Help*) – 4.95

Group M: middle scores for literary quality – mean score

- Jenna Blum, *Het familieportret* (*Those Who Save Us*) – 4.90
- Hannah Shah, *De dochter van de imam* (*The Imam's Daughter*) – 4.89
- Tatiana de Rosnay, *Het huis waar jij van hield* (*The House I Loved*) – 4.89
- Anita Amirrezvani, *Dochter van Isfahan* (*The Blood of Flowers*) – 4.89
- Tatiana de Rosnay, *Het appartement* (**The Neighbour*) – 4.82
- Tatiana de Rosnay, *Kwetsbaar* (*Moka*) – 4.81
- Sarah McCoy, *De bakkersdochter* (*The Baker's Daughter*) – 4.77
- Amanda Hodgkinson, *Britannia Road 22* (*22 Britannia Road*) – 4.77
- Jenna Blum, *In tweestrijd* (*The Stormchasers*) – 4.71
- Margaux Fragoso, *Tijger, tijger* (*Tiger, Tiger*) – 4.71
- Tatiana de Rosnay, *Haar naam was Sarah* (*Sarah's Key*) – 4.67
- Tatiana de Rosnay, *Die laatste zomer* (*A Secret Kept*) – 4.59
- Sarah Blake, *De laatste brief* (*The Postmistress*) – 4.59
- Nicci Gerrard, *Het weerzien* (*The Middle Place*) – 4.53
- Janelle Brown, *Alles wat wij wilden was alles* (*All We Ever Wanted Was Everything*) – 4.23
- Jodi Picoult, *Negentien minuten* (*Nineteen Minutes*) – 4.17

Group L: lowest scores for literary quality – mean score

- Joy Fielding, *Roerloos (Still Life)* – 3.88
 Lynn Austin, *Eindelijk thuis (Until We Reach Home)* – 3.85
 Elizabeth Gilbert, *Toewijding (Committed)* – 3.80
 Jean M. Auel, *Het lied van de grotten (The Land of Painted Caves)* – 3.66
 Monika Peetz, *De dinsdagvrouwen (*Tuesday Women)* – 3.59
 Elizabeth Gilbert, *Eten, bidden, beminnen (Eat, Pray, Love)* – 3.51
 Karen Kingsbury, *Laatste dans (A Time to Dance)* – 3.50
 Emily Barr, *Klasgenoten (Out Of My Depth)* – 3.49
 Karen Kingsbury, *Nooit te laat (Redemption)* – 3.47

As in Chapter 3, I used the Stylo package for R (Eder et al. 2016). My first experiment was a principal components analysis (PCA) of the thousand most frequent words in all 34 novels together. In the visualization of the measurement in Figure 4.3, each group has been given its own grey scale, to make it easier to spot the major differences. More about the method can be found in Chapter 3. In reading the graphs, the essence is that the closer books are positioned to each other, the more similar they are in their frequency distribution of the thousand most frequent words. Figure 4.3 shows that the three groups overlap by a wide margin. This suggests that word usage cannot very well explain the differences in literary appreciation. It is noteworthy that the three Spanish novels are collectively positioned in the upper left quadrant, quite far away from the other novels receiving the highest scores for literary quality, which suggests to me that their shared Spanish origin is reflected in the style of the translations.

Even when I selected other amounts of most frequent words or applied other kinds of measurements, the same novels ended up as relative outliers. In the group with the highest literary quality Emma Donoghue's *Kamer (Room)* stood somewhat apart; I will examine this novel in more detail in the next section. In the middle group Sarah Blake's *De laatste brief (The Postmistress)* was an outlier, and in the group with the lowest literary quality both titles by Elizabeth Gilbert, *Toewijding (Committed)* together with the book it was a sequel to, *Eten, bidden, beminnen (Eat, Pray, Love)*, were outside of the core group. *Toewijding*, by the way, received a slightly higher average score for literary quality than *Eten, bidden, beminnen*: 3.80 as opposed to 3.51. I suspect that this was because those readers who gave *Eten, bidden, beminnen* very low scores did not read the sequel *Toewijding*, so that the mean scores based on the opinions of the repeat readers was slightly higher – the same pattern as described in Chapter 3 for the three volumes of the *Fifty Shades* trilogy, where the second and third volumes received slightly higher scores for literary quality than the first one.

As in Chapter 3, I used the keyword function in Laurence Anthony's AntConc software to find out which words occurred significantly more often than could

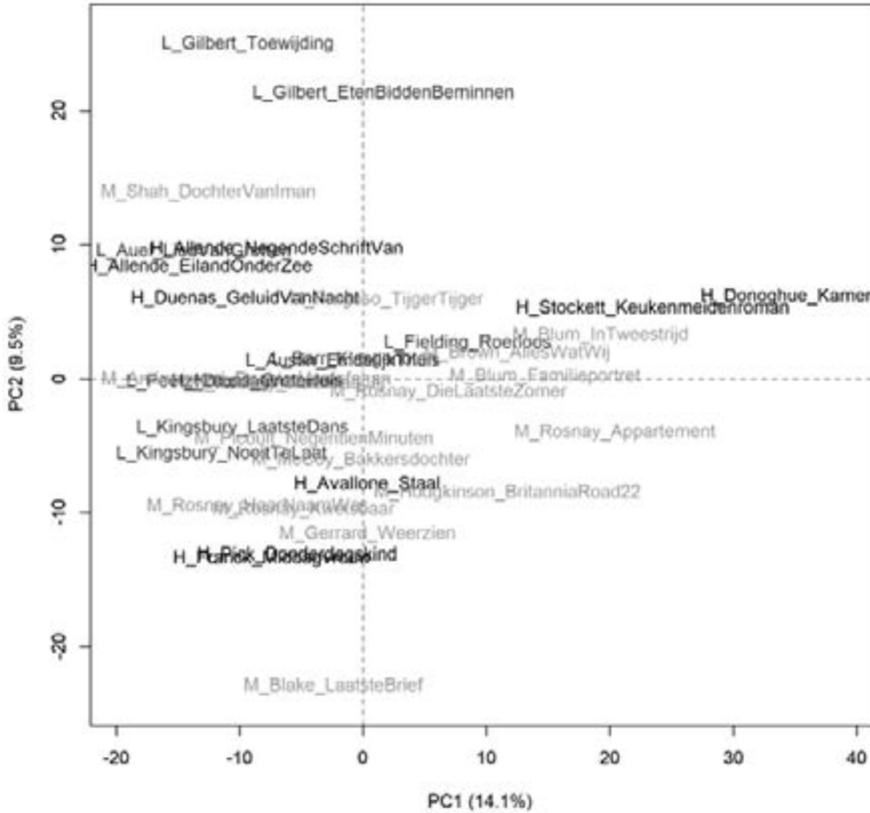


Figure 4.3: Translated Literary novels written by female authors, PCA (1000 most frequent words). Scores for literary quality: H_ (highest), L_ (lowest), and M_ (middle group). Nicole Krauss, *Het grote huis* (*Great House*) hides among other books on the zero line to the left of the intersection (which in a way represents the average of the 34 novels).

be statistically expected. For this, I left out the middle group and only compared the nine novels that readers found most literary with the nine novels that received the lowest average score for literary quality. This resulted in the observation that the novels with the highest literary quality contained relatively more personal pronouns indicating a first-person narrative (*ik, mijn, me* – I, my, me), and more auxiliary verbs in the present tense (*zeg, zegt, kijk, kijkt* – say, says, look, looks) than the novels that were awarded the lowest literary quality. Conversely, the nine novels with the lowest literary quality rating contained more personal pronouns *ze* (she/they) and *haar* (her), which may indicate a predominance of third-person narratives and possibly also a greater representation of female characters, but also more *je* (you/your, singular) and *jullie* (you/your, plural), which may indicate more dialogue. Furthermore, they showed many more past tense forms here: *zei,*

vroeg, keek (said, asked, looked). In the novels with low scores for literary quality we also see a larger share of the conjunctions *toen* (then) and *dat* (that; the use of some conjunctions was discussed in Chapter 3). The text of the lowly rated *Eten, bidden, beminnen* (*Eat, Pray, Love*) agreed in more ways than one with the highest scoring novels.

Two words that were characteristic for the novels with the highest scores can be traced back to the book that was a bit on the fringe in Figure 4.3: *Kamer* (*Room*) by Emma Donoghue. They are *ouwe* (old), which invariably precedes the male personal name *Nick* in the novel, and *wc* (wc, toilet). The word *wc* occurs a total of ninety times in *Kamer*, which is a remarkably high number in the entire corpus of 401 books. Except for *Een keukenmeidenroman* (*The Help*) by Kathryn Stockett, where *wc* is found forty times, the other novels hardly refer to it. In the previous chapter, *Kamer* also ended up as an outlier in the use of the conjunction *hoewel* (although). In the next section I will look into this novel in more detail.

Donoghue, *Kamer* (*Room*)

Emma Donoghue's novel *Room* was published in 2010. The Dutch translation *Kamer* by Manon Smits appeared on the market in the same year. The style of the Dutch version very closely recreates that of the original English text. To illustrate, I quote the opening sentences of the English version:

Today I'm five. I was four last night going to sleep in Wardrobe, but when I wake up in Bed in the dark I'm changed to five, abracadabra. Before that I was three, then two, then one, then zero. 'Was I minus numbers?'

'Hmm?' Ma does a big stretch.

'Up in Heaven. Was I minus one, minus two, minus three – ?'

'Nah, the numbers didn't start till you zoomed down.'

'Through Skylight. You were all sad till I happened in your tummy.'

We are looking through the eyes of Jack, a five-year-old boy who all his life has never left the room in which he was born. His mother, the reader soon realizes, is kept prisoner in what looks like a small wooden shed. The shed has no windows except a skylight and is almost completely soundproof. Jack's mother lived alone in the room for a very long time before Jack arrived. In Jack's words and thoughts:

'Then you wished and wished on your egg till you got fat.'

She grins. 'I could feel you kicking.'

‘What was I kicking?’

‘Me, of course.’

I always laugh at that bit.

‘From the inside, *boom boom*.’ Ma lifts her sleep T-shirt and makes her tummy jump. ‘I thought, *Jack’s on his way*. First thing in the morning, you slid out onto the rug with your eyes wide open.’

I look down at Rug with her red and brown and black all zigging around each other. There’s the stain I spilled by mistake getting born.

Jack’s short history is reflected in the language he uses: many nouns are capitalized and are not preceded by an article or a pronoun. Jack speaks of ‘Wardrobe’, ‘Bed’, ‘Skylight’, and ‘Rug’ whereas his mother speaks of ‘the rug’ in the quotation above. For Jack, these words are proper names: they refer to an object of which there is only one in his closed world; for us, proper names usually refer to one specific person, object or place, to distinguish them from other persons, objects, or places – which may have the same name, by the way: there may be more than one boy of five called Jack, for instance (more about names in novels in Van Dalen-Oskam 2013). This is maintained throughout the novel, which makes the book stylistically very remarkable.

Two participants in The National Reader Survey who awarded the book the maximum score of 7 for both literary and general quality (how good the book was perceived to be) wrote: ‘Wonderful perspective, original use of language and captivating story’ (ID_12992). ‘Special book with intriguing theme. But the form is also beautiful, told entirely from the perspective of the locked-up child. Also, the child’s language was very appropriate’ (ID_10806). A participant who gave the novel a relatively low 4 for literary quality had trouble with the language: ‘Suspenseful, language in function of the child-character inventive, but because of that not “normal”, “correct” Dutch’ (ID_6013; this reader did give it a 6 for general quality). Only one of the respondents explicitly indicated that she did not like the style: ‘Characters had little depth. The whole writing style irritated me, although the story had “suspense” and “came together” in the end. It *was* clever how that small room was described’ (ID_4742). She gave the book score 2 for literary quality and a 3 for general quality.

A visualization of all the ratings shows a very different picture than the one presented earlier in this chapter for *Eten, bidden, beminnen* (*Eat, Pray, Love*). Of the 534 readers (526 of whom were women) who rated *Room*, only a few gave very low scores (Figure 4.4).

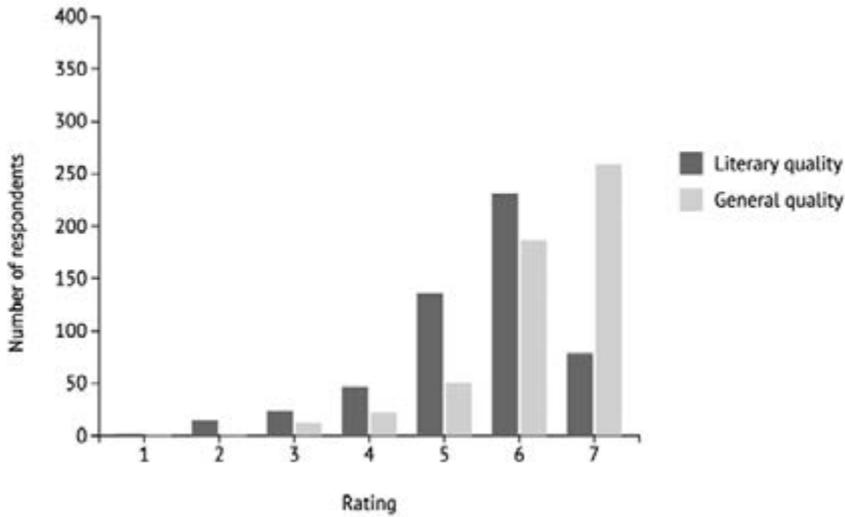


Figure 4.4: Ratings of Emma Donoghue, *Kamer (Room)*

Kamer (Room) was rated with significantly higher scores for literary quality than *Eten, bidden, beminnen (Eat, Pray, Love)*. In the earlier visualization of a PCA based on the thousand most frequently used words (Figure 4.3), both books had a place at the sides, but at different ends of the visualization. This suggests they are both exceptional in relation to the other novels, but in different ways. It is unlikely that their different position is solely due to a difference in literary quality. It is very probable that the exceptional perspective Donoghue applied was responsible for the novel's place in relation to the other translations. All the missing articles and pronouns in little Jack's clauses may already have done the trick to set the novel apart.

The main observations from the first part of this chapter are that the translated novels that are considered most literary contained unexpectedly more occurrences of words indicating a first-person narrative in the present tense. The novels awarded the lowest literary quality had more third-person narration, more female characters, and possibly more dialogue, in the past tense. The use of some conjunctions in the lowest-rated novels may indicate a slightly more explanatory style on the sentence level. Furthermore, the style of Gilbert's two novels seemed to be more literary than the participants in The National Reader Survey expressed in their scores and motivations, and the novel *Kamer (Room)* most likely differed measurably from the other books in the group because of the very specific child's perspective, reflected in the writing style. In the remainder of this chapter I will discuss the literary novels written in Dutch by female authors.

The Dutch Perspective

In addition to the 34 translated Literary novels by female authors the research corpus contains 23 novels written in Dutch. I will provide some more information about these novels, indicate how the participants in The National Reader Survey rated them, and report on what showed up in text analysis experiments. Again, I divided the novels into three groups: those with the highest scores for literary quality, those with the lowest scores, and a middle group.

Group H: highest scores for literary quality – mean score

- Margriet de Moor, *De schilder en het meisje* (**The Painter and the Girl*) – 5.90
- Marente de Moor, *De Nederlandse maagd* (*The Dutch Maiden*) – 5.86
- Anna Enquist, *De verdoovers* (**The Anaesthetists*) – 5.81
- Tessa de Loo, *De grote moeder*† (**The Great Mother*) – 5.67
- Jessica Durlacher, *De held* (**The Hero*) – 5.66
- Renate Dorrestein, *De stiefmoeder* (**The Stepmother*) – 5.53
- Renate Dorrestein, *De zondagmiddagauto*† (**The Sunday Afternoon Car*) – 5.49
- Franca Treur, *Dorsvloer vol confetti* (**Confetti on the Threshing Floor*) – 5.36

Group M: middle scores for literary quality – mean score

- Vonne van der Meer, *De vrouw met de sleutel* (**The Woman with the Key*) – 5.35
- Joshua Zwaan, *Parnassia* (**Parnassia*) – 5.11
- Vonne van der Meer, *Het zingen, het water, de peen*† (**Carolling, Water, Carrot*) – 5.02
- Renate Dorrestein, *De leesclub* (**The Reading Group*) – 4.98
- Susan Smit, *Vloed* (**High Tide*) – 4.96
- Simone van der Vlugt, *Jacoba, Dochter van Holland* (**Jacoba, Daughter of Holland*) – 4.86
- Simone van der Vlugt, *Rode sneeuw in december* (**Red Snow in December*) – 4.84
- Esther Verhoef, *Tegenlicht* (**Backlight*) – 4.20

Group L: lowest scores for literary quality – mean score

- Rita Spijker, *Tussen zussen* (**Between Sisters*) – 4.06
- Loes den Hollander, *Het scherventapijt* (**A Carpet of Shards*) – 3.92
- Naima El Bezaz, *Vinexvrouwen* (**Middle-Class Women in Modern Suburbia*) – 3.54
- Heleen van Royen, *Sabine*† (**Sabine*) – 3.36
- Heleen van Royen, *De mannentester* (**The Tester of Men*) – 3.18
- Van Sambeek, *Koninginnenrit* (**Queen's Ride*) – 3.12
- Liza van Sambeek, *Bloed zaad en tranen* (**Blood, Seed, and Tears*) – 3.05

The four titles marked with a cross are stories that were published separately as a small booklet. I will not include them in the computer analyses because they differ too much in length from the novels and for that reason often end up as outliers without really telling us anything useful about their writing style and word usage. One of them was written by Renate Dorrestein – more about her in the next section of this chapter. The story was first published in 2012, but Dorrestein stated in the preface that she wrote it in the early 1980s. She had completely forgotten that she had written it:

Surprised, I read it in one sitting. My thirty years younger self hadn't economized much on adjectives and adverbs, and the narrative pace could have been faster, but the brooding atmosphere of the wicked fairy tale enchanted me, as did the heavy-handed and outlandish characters (Dorrestein 2012: 25, my translation, cf. Van Dalen-Oskam [2021] 148–149).

Dorrestein did not mention whether she refined the text before publication, and therefore we do not know whether she resisted the temptation to delete a few adverbs and adjectives and if this story measurably differs in this respect from her more recent work. I will not go into this any further in this monograph although I will come back to adjectives and adverbs in Chapter 6 of this book.

The novel that received the highest literary quality was Margriet de Moor's *De schilder en het meisje* (*The Painter and the Girl*). It is ranked 36th on the list of 401 novels ordered from highest literary quality to lowest, four places below the highest ranked (translated) female author in the list, Nicole Krauss. The book was first published in 2010 and was rated by 611 participants. The novel describes both a day in the life of the famous Dutch painter Rembrandt in May 1664, and the life story of a Danish immigrant, Elsje Christiaens, who is executed on Dam Square that day for murdering her landlady. Rembrandt does not attend the execution but his son does. The painter is so intrigued by his son's account of the event that he sets off with his drawing materials to the gallows field in the late afternoon to draw Elsje. Rembrandt's day runs chronologically from just before eleven in the morning until the end of the afternoon, but Margriet de Moor narrates a great deal about his life up to that time, especially about the death of his second wife a year earlier. He is still overwhelmed with grief about her passing. The life of Elsje Christiaens is told in bits and pieces, sometimes jumping backwards and forwards in time. Two existing drawings by Rembrandt of a young woman on the gallows prompted De Moor to give a virtually unknown young woman a whole (fictional) life.

In March 2022, the translation database of the Letterenfonds (Dutch Foundation for Literature) indicated Margriet de Moor's work had been translated into 26 languages. A total of eighteen titles was available in German, nine in Czech, and

seven in French and English. *De schilder en het meisje* has not been translated into English yet. For an impression of the writing style, I present my own provisional translation of a passage from chapter 3 of the novel (cf. Van Dalen-Oskam [2021] 150).

He walked on, down the Achterburgwal, his thoughts both on his rejected painting and his dead wife. He turned into Korte Kolksteeg, took the bridge at the end and crossed the Nieuwezijds Voorburgwal. Of course she hadn't wanted to die. If anyone had been 100 per cent right about that, it would have been her. Optimism is one of the side-effects of being unsuspecting. Even when, at the beginning of summer, it began to smell of the foul plague here in the city, she did not immediately notice it. A person who is not attuned to decay does not have the best nose for the decay that surrounds them. She was good, beautiful, her body very soft. Her body was what she was. If ever, at the pawnshop or at an auction, a bust of a woman with her neck, shoulders and breasts were offered, such a beautiful piece of creamy Dolomite alabaster whose surface is so soft to the touch like powder, he would buy it at once and wrap a white fur around it at home. The fact is that plague often does not smell like plague at all in the beginning, not heavy and dirty like an old piece of boiled meat thrown in the canal, but spicy, like a sea breeze.

There were still many people walking in the street, mostly in a bit of a hurry, heading in the opposite direction towards Dam Square. They would not get a place at the front anymore. There was the Nieuwe Dijk with the bridge over the Damrak on the right around the corner.

The painter, a well-known figure in the city, stared languidly into the distance, not wanting to cross his thoughts with those of the others by saluting.

Most of the readers who reviewed the book were very impressed with it. The average score for literary quality was 5.90 and for general quality 5.89. Of the 611 readers who shared their opinion, 41 motivated their rating, 29 women and 12 men. This reader gave the book the highest score for literary quality, 7, and a 6 for general quality: 'The sentences were well formulated and came across to me as a work of art. I also noticed that the story was not so much about events and the build-up of suspense, but more about how a certain event or encounter can affect people. The author did not linger on the narrative level, but went a little deeper into things' (ID_3988). The next reader rated the literary quality a 6 and the general quality a 4: 'Because the reader had to do a lot of thinking and everything was not presented on a silver platter by the writer' (ID_10959).

I now turn to one of the lowest scoring Dutch Literary novels: *De mannentester* (**The Tester of Men*) written by Heleen van Royen. Victoria Kramer is a young woman with a very special job: she rents herself out to seduce men who are suspected of infidelity by their wives or girlfriends. Victoria is always successful, until she

decides to put her mother's new boyfriend to the test. In the literary ranking, the book ended up at place 354 out of 401 in terms of literary quality, with an average score of 3.18 (Fig. 4.5). Of the 216 readers who cast their votes, 22 justified their scores. One of the six men who did so wrote: 'I found it mainly focused, in its violent and erotic details, on sensation' (ID_6829). He gave the book a 2 on the scale of 7 for both literary quality and general quality. Of the sixteen female readers who substantiated their scores, one scored the book a 1 on both aspects: 'Because this book is known purely because of the author's name, had it been written by any other writer, publishers would have put it in the shredder immediately. Furthermore, *De manentester* is purely superficial, too much nonsense' (ID_2668). Another woman also gave it a 1 for literary quality, but a 5 for general quality: 'With literature, I expect depth (not necessarily complexity, symbolism, beautiful/surprising ways of describing). *De manentester* is a fascinating story. Van Royen does have humour' (ID_4800). A final striking opinion I would like to quote takes us back to Chapter 3, because this (female) reader wanted to place the novel outside the category of the Literary novel: 'It is a chick lit novel. It does not contain any great questions about life or profound themes. It has no other purpose than common entertainment' (ID_8913). This reader's motivation reflected the by now familiar conception: what makes a novel literary is its writing style and its deeper layers and that the protagonists experience personal growth. We will have to take a closer look at the novels themselves to find out which linguistic features played a role in the formation of such widely varying judgements.

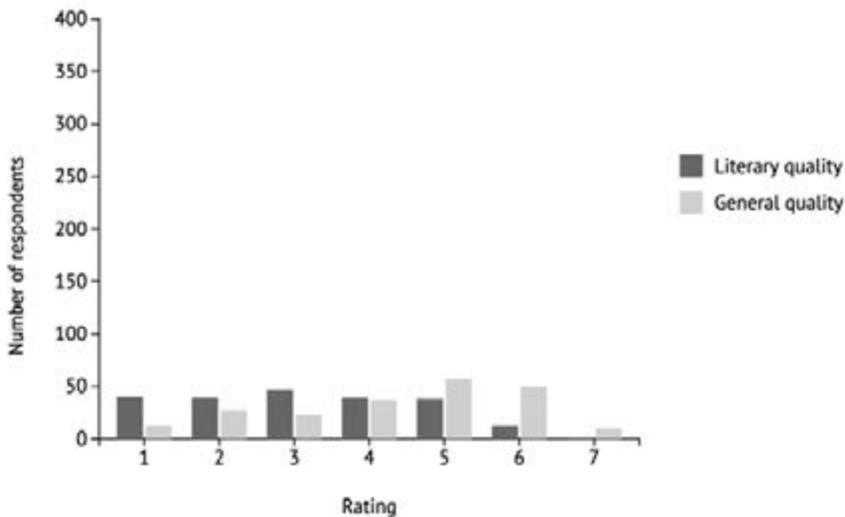


Figure 4.5: Ratings of Heleen van Royen, *De manentester* (**The Tester of Men*)

As for the translated novels I started with a PCA based on the thousand most frequent words from all nineteen Dutch novels (Figure 4.6). To make it easier to spot the titles from the three groups (highest and lowest literary quality and the middle group), they have each been given a different shade of grey in the visualization. I have explained PCA in more detail in Chapter 3. As with the translated novels, the three groups of Dutch novels with different score ranges largely overlap, which means that differences in literary quality cannot be explained by the word frequencies that were examined.

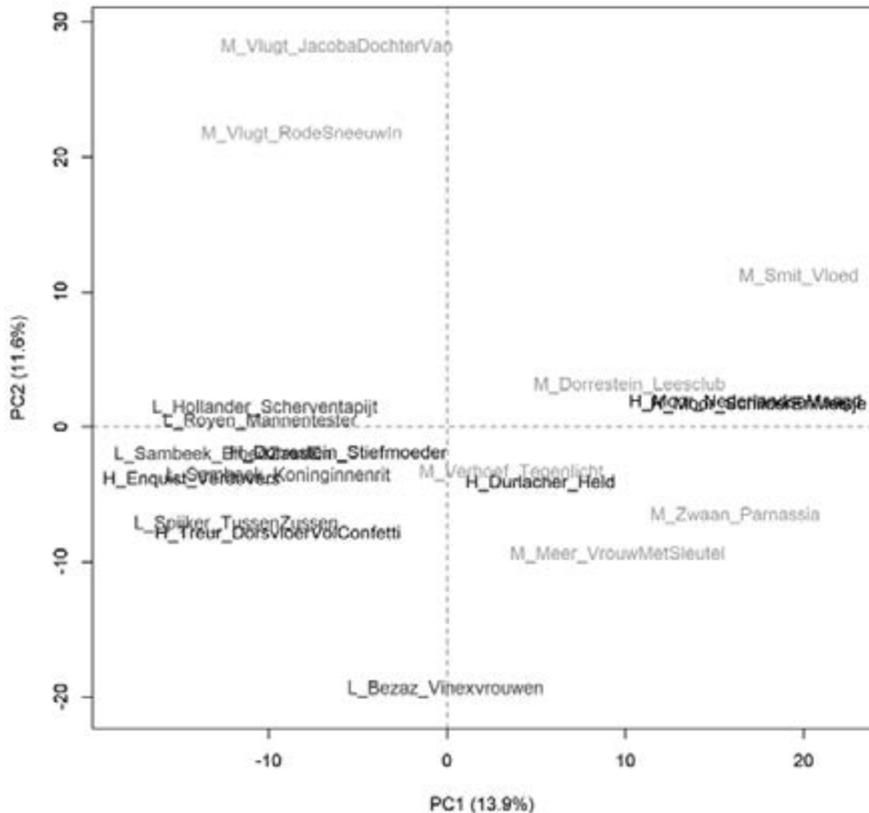


Figure 4.6: Dutch Literary novels written by female authors, PCA (1000 most frequent words).

Scores for literary quality: H_ (highest), L_ (lowest), and M_ (middle group).

The two almost completely overlapping novels below Dorrestein's *De leesclub* are Margriet de Moor, *De schilder en het meisje* (**The Painter and the Girl*) and Marente de Moor, *De Nederlandse maagd* (*The Dutch Maiden*), the two novels with the highest mean scores for literary quality in this selection.

Figure 4.6 begs several questions about the relative outliers in the visualization. At the bottom of the graph we find *Vinexvrouwen* by Naima El Bezaz (1974–2020). In 38 short chapters, the book describes life in a ‘Vinex neighbourhood’ – a new neighbourhood with uniform and affordable housing intended for young families – in Almere, a satellite city for former inhabitants of Amsterdam. The first-person narrator Naima, her husband, and their two daughters are the first family of Moroccan descent in the neighbourhood, and this causes some unrest with the neighbours, but also with Naima herself. The book is not available in English; a literal translation of the title would be ‘Vinex Women’ but this will probably not be helpful to those not familiar with the Dutch Vinex concept. My provisional translation **Middle-Class Women in Modern Suburbia* should be seen as a very rough approximation. In March 2022, the translation database of the Letterenfonds (Dutch Foundation for Literature) did not mention any translations of any work of Naima El Bezaz yet, so below is a tentative translation of a quote from chapter 24 (cf. Van Dalen-Oskam [2021] 162–163).

It was Easter. And for the first time, I went to the Protestant church to attend the Easter celebration organized by my daughter’s school. I was as excited as my eldest and I wanted everything to be perfect. Muslim women wear headscarves when they go to the mosque. The Queen and others wear hats in church, I always saw that on TV when there was another wedding or christening. But I didn’t have a hat. My medicine supplier had some, but when we were trying on hats in her bedroom, we discovered that the circumference of our heads didn’t quite match. I needed something smaller, but where could I find a decent hat? My daughter admired herself in the mirror because she wanted to look beautiful, as she thought she should. The Christian school had clear rules, good standards, and values, so my eldest taught me a lot. Which words to use and which not. But now the hat. I found a number of them on Otto.nl. Some looked like the ones I saw on TV in a church. I ordered one and it would be delivered within a few days.

The book, published as a novel, has obvious autobiographical traits and may be characterized as auto-fiction. The main character has the same name as the author, and the events described seem to have a high degree of realism, but how fictional they are remains unclear. The book was controversial and sparked outrage among readers. This is reflected in the motivations we received from participants in The National Reader Survey. One male reader who rated the book with a 3 for literary quality and a 6 for general quality wrote: ‘Because it’s a daily account of life there, without frills’ (ID_10043). Another male reader gave the book a 1 on both scales, stating: ‘No depth, stylistically very mediocre, content uninteresting. At the level of chit-chat’ (ID_12219). And to also give a voice to two female readers: a participant who gave the book a 1 for

literary quality and a 4 for general quality wrote that '[i]t is a light-hearted book, you don't have to make any effort for it at all. But it is a nice book' (ID_7103). Several women would rather classify the novel as belonging to another genre, including the following reader who gave it a 3 on both accounts: 'In my opinion, the book had no deeper meaning. Besides, the book read more like a chick lit than a literary work' (ID_3461).

It is difficult to find out why this novel is a relative outlier compared with the other eighteen books in this group of novels. My hunch is that it may have to do with the auto-fictional nature of the book, and therefore that genre (or subgenre) plays a role here. Because there are too few comparable novels in the corpus, this possibility cannot be further explored.

At the top of the graph in Figure 4.6, as relative outliers, sit two novels by Simone van der Vlugt that may be characterized as historical novels. Here again, one might think that genre or subgenre – in this case the historical novel – may be responsible for differences in style that are reflected in vocabulary differences. However, there are several other historical novels in the selection: Susan Smit, *Vloed* (**High Tide*), Josha Zwaan, *Parnassia* (**Parnassia*), and also the two highest scoring novels by Margriet de Moor and Marente de Moor, the two novels that mostly overlap in the graph just below Renate Dorrestein's definitely not historical novel *De leesclub* (**The Reading Group*). This exception calls the genre explanation into question. Dorrestein is represented with two novels in this selection, which end up in different parts of the graph – a bit comparable to what we saw for the two literary novels written by Mankell in Chapter 3. I will go into this in more detail in the next section of this chapter.

What is also intriguing in the graph is the closeness of the novels by Margriet de Moor and by Marente de Moor. I have already introduced Margriet de Moor's work, with special focus on *De schilder en het meisje* (**The Painter and the Girl*), and presented some motivations for its ratings. I will now do the same for Marente de Moor's work. So far, she has published six books, of which *De Nederlandse maagd* was the third. As of March 2022, Marente de Moor's work had been translated into sixteen different languages; four of her titles were available in German and two in Bulgarian, French, Italian, and Spanish. *De Nederlandse maagd* was the only novel in the set of Dutch Literary novels written by female authors that was available in an English translation.

De Nederlandse maagd appeared in Dutch in 2010 and was awarded the prestigious Dutch AKO Literature Prize in 2011 and the European Union Prize for Literature in 2014. The English translation by David Doherty, *The Dutch Maiden*, was published in 2016. Eighteen-year-old Janna leaves Maastricht by train in September 1936 to stay with an old friend of her father's for a few months. Egon Von Bötticher lives in a village near Aachen and is a renowned fencing instructor (maître). Janna's father is a doctor who met Von Bötticher in 1914 when he worked for the Red Cross as a

medical aid worker on the battlefields just across the Dutch border. For reasons that are unclear to Janna, the relationship between the two men has long been strained. Janna came into contact with fencing at the 1928 Olympic Games in Amsterdam and has been enchanted by the sport ever since. Her father eventually gives her the opportunity to continue her studies at Von Bötticher's, together with the beautiful identical twins Friedrich and Siegbert. Janna is completely fascinated by their instructor and his mysterious past, but the political situation in Germany in 1936 brings her stay in the house to a fateful end. Below I quote a longer excerpt from chapter 9 of Part 1 of the English translation, which carefully reflects the style of the original Dutch (cf. Van Dalen-Oskam [2021] 160).

He [Von Bötticher] picked up a foil and stood in front of the mirror, poised to attack. As I had expected, his stance was perfect. All of the power in his tensed body was channelled into his elegant left hand, held effortlessly in mid-air as if his right held no weapon at all. Between the tip of his foil and the mirror there was no more than a breath, just enough to suggest that his reflection was an opponent of flesh and blood. A fraction closer and the illusion would have been shattered.

'Two identical sabreurs, but trained differently. Your opponent becomes rattled: which one is he facing? He tries to respond exactly as he responded to your likeness, but the two of you fence differently every time. This unsettles him. Is this the sabreur who repeatedly braves his attacks, parrying nimbly, or is it the other, the one who always seems to be on the retreat...' – without altering his stance he backed off at lightning speed – '... only to lunge forward and make good his position? Racked by doubt, he concedes a hit.'

It struck me as an absurd plan. Didn't all fencers look pretty much alike when they were suited and masked? Besides, the names were announced before every match and opponents on both sides were changed at the same time. Nevertheless the twins were delighted.

'Like Zorro!' Friedrich cried. 'No one knows Zorro's true identity either!'

Von Bötticher frowned. He had no idea what the boy was talking about.

Of the 629 readers who rated the book, 36 wrote a motivation for their score. A female reader gave the book the highest score, a 7 for both literary quality and general quality: 'The style of writing, the sentence structure and the vocabulary. Also won a number of literary prizes, I believe' (ID_1398). Another female participant gave it a 6 on both scales: 'Unexpected, unusual theme, excellent writing style' (ID_10970). A few people were considerably less positive; the following reader rated both literary as general quality with a low 2 and justified her decision as follows: 'The style was stiff and long-winded. I found the book unreadable' (ID_1201). The next extensive motivation was from a female reader who gave the book a 7 for literary quality and

a 6 for general quality: ‘You experience the development of the main character: you get to know her thoughts and feelings. Her doubts, but also her curiosity about the fencing teacher hold your attention. You increasingly feel the threat of an unrequited love. The book is written in beautiful Dutch: varied use of words, original way of phrasing. A language that is too beautiful for everyday life’ (ID_13197).

In the visualization of the PCA based on the thousand most frequent words in Figure 4.6 Marente de Moor’s *De Nederlandse maagd* (*The Dutch Maiden*) seemed to completely overlap with Margriet de Moor’s *De schilder en het meisje* (**The Painter and the Girl*). One could hypothesize that the close relationship between the authors may have played a role: Marente is Margriet’s daughter. The Polish computational literary scholar Jan Rybicki has shown that the works of the English Brontë sisters always cluster together in similar stylistic analyses of nineteenth-century English literary fiction (Rybicki 235–236). When asked to explain this clustering during workshops he teaches on the Stylo package in R, Rybicki usually suggests this may have to do with the seclusion of the family in which the sisters grew up and shaped their writing careers. In 2021, this idea was confirmed by Rachel McCarthy and James O’Sullivan. To test whether the novels by Margriet de Moor and Marente de Moor resemble each other in diction because the authors have spent a lot of time in each other’s presence and have been influenced by each other in their work, I applied a cluster analysis in addition to the PCA. I used the Stylo Package for R and used Classic Delta as a measure; in the visualizations I made, the Delta score is shown on the horizontal axis. The statistical approach of the cluster analysis establishes pairs of most similar texts or clusters based on words and word frequencies. The further to the right in the graph two books are linked, the more similar they are in the frequencies of the words analysed. Cluster analyses based on different amounts of most frequent words presented a result that aligned with what Figure 4.7 showed for the thousand most frequent words. However, in this different statistical approach compared to PCA, the novels of Marente de Moor and Margriet de Moor no longer closely resembled one another, but each clustered with one of the other novels. Still, the four of them were the most similar. Marente de Moor’s *De Nederlandse maagd* (*The Dutch Maiden*) most closely resembled another historical novel, Susan Smit’s *Vloed* (**High Tide*). Together they cluster with the pair consisting of Margriet de Moor’s historical novel *De schilder en het meisje* (**The Painter and the Girl*) and Renate Dorrestein’s *De leesclub* (**The Reading Group*), the latter of which is definitely not a historical novel.

What we can again conclude from this experiment is that genre is not a sufficient explanation for stylistic similarity. It is quite possible that the age of the authors also plays a role. Marente de Moor (born in 1972) and Susan Smit (born in 1974) are almost contemporaries. The other two women differ more in age: Margriet de Moor was born in 1941 and Renate Dorrestein in 1954. Whether age and generational

differences do indeed influence writing style can only be confirmed when we conduct much more extensive research based on a much larger corpus of novels, so an answer to this question will not be available for now – and the same goes for the question if the family relation between Margriet de Moor and Marente de Moor has influenced similarities in their writing style.

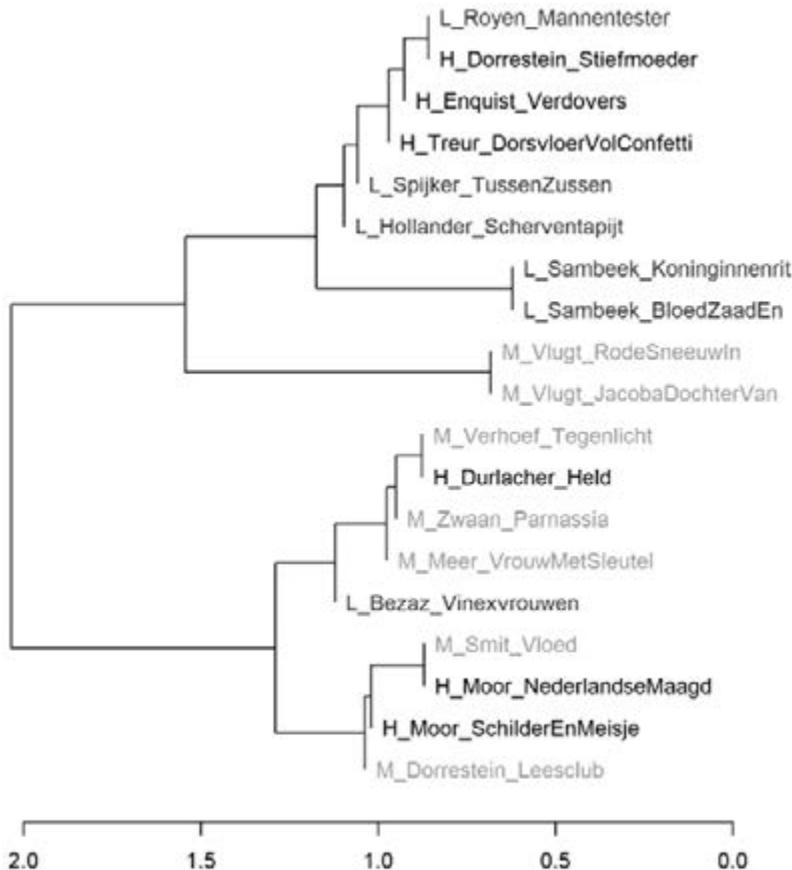


Figure 4.7: Dutch Literary novels written by female authors, Cluster Analysis (1000 most frequent words). Scores for literary quality: H_ (highest), L_ (lowest), and M_ (middle group).

Dorrestein in More Detail

Renate Dorrestein (1954–2018) is mainly known as the author of novels but she was also active as a speaker and as a teacher of creative writing. Some of her non-fiction

publications dealt with writing, sharing her writing methods, and suggesting best practices to aspiring authors. Several of her novels are translated. According to the translation database of the Letterenfonds (Dutch Foundation for Literature), the number of receiving languages as of March 2022 was seventeen. German topped the list with a total of fourteen titles, followed by Italian with eight, French and English both with five and Finnish and Swedish with three. The two Literary novels in our research corpus were not among the five novels translated into English. *De stiefmoeder* (**The Stepmother*) is available in German and in Finnish; in the list of 401 novels ranked for literary quality, in our research corpus, it occupied the 64th rank with a mean score of 5.53. *De leesclub* (**The Reading Group*) had not been translated; it can be found at rank 99 with a mean score of 4.98 for literary quality. In the PCA and the cluster analyses reported on and visualized in the previous section, *De leesclub*, published in 2010, invariably ended up in a different cluster than *De stiefmoeder*, published in 2011 (Figures 4.6 and 4.7). To explore their differences I will provide a very short summary of the two novels, select some motivations of readers' opinions, and provide a very provisional translation of my own hand of a somewhat longer quote from the novels.

De leesclub (**The Reading Group*) is presented as a court hearing of which we only get the oral report of a reading group that went on a boat trip in Scotland with the Dutch writer Gideon de Wit. The reading group consists of seven middle-aged women who read a truly heroic number of books and always prepare their meetings with authors down to the last detail. The seven members Tillie, Jo (short for Johanna), Willemien, Annabel, Martha, Leonie, and Barbara together act as narrator, who always speaks of 'we' and 'us'. The written account by this first-person plural narrator is interspersed with 45 footnotes of tourist information for the reader travelling in Scotland, compiled by Leonie. De Wit is the author of 'one of the most talked-about oeuvres of this century and the last', and is highly regarded and showered with awards, especially outside of the Netherlands. The women chose him as their tour guide for a boat trip themed 'In the spirit of Moby Dick'. But that trip does not end well – the very first sentence of the novel suggests that the reading group ended up murdering de Wit. An important link in the chain of events is the following scene in the sixth chapter, in my translation (cf. Van Dalen-Oskam [2021] 167–168):

[B]efore he [Gideon de Wit] embarked on his duties as a tour guide, he wanted to clear up a misunderstanding that had haunted him for years. He cast a meaningful look at each and every one of us. A misunderstanding that, according to some quacks, was even the main reason for his conversion to Islam: he would have claimed on several occasions that women could not write. Pure nonsense and slander. Such a statement was semantically impossible. In reality, he had only

occasionally meekly suggested that in order to write well, one had to have balls. It was easy to spot the work of female authors. Books about trifles. Books full of nitty-gritty details. Books without weight and mass. Books without balls.

That touches a nerve.

The blood went to our heads. We had to bite our lips. An old nag. But perhaps it was just a question of taste, or perhaps a literary judgement was being passed that we as amateurs could never match. Nevertheless, our heart went out to the female authors we had met over the years in our series of literary encounters. We had devoured many of their ball-less books, they had sometimes changed our lives or at least shed new light on them. We had often found them to be wise women with humour and great psychological insight. We decided we had to stand up for them now. We had to defend the weight and mass of their work in no uncertain terms. We could not deny them without feeling indefinitely guilty, and didn't guilt lead to penance? But would Gideon de Wit still take us seriously if we started talking about our beloved female writers?'

The designation 'anti-roman' (anti-novel) did not appear on the title page of the first print of *De leesclub*, but this is the label Dorrestein herself retrospectively attributed to the novel in the overview of her previously published works in her later books. She did not use that for any of her other works, as can be seen in the front matter of her 2017 book about writing that I will go into shortly.

Of the sixteen motivations we collected about *De leesclub* in The National Reader Survey, I will quote two. This female reader gave the novel the maximum score of 7 for both general and literary quality: 'A thin book with a lot of content. I read it twice in a row: the first time for the story, the second time for the rest. It leaves a lot of room for personal interpretation and imagination, not everything is told' (ID_303). A completely opposite opinion can be found in the motivation of another reader: 'Thought it such a stupid book and literary means that it has interesting layers and hands the reader something to take away. Humourless, boring, and totally uninteresting, I found it. I'm not a Renate Dorrestein fan' (ID_2461). She gave the book a 1 for general quality and a 2 for literary quality.

In *De stiefmoeder* (**The Stepmother*) Claire Paagman goes on a solo trip for the first time in years. At the last minute, after a heated argument, her husband Axel does not come along. She travels to the English coast for an exhibition of her colourful quilts. The exhibition is organized by the Royal Quilting Academy, which has awarded Claire a prize that she has to collect with a speech. Shortly after arriving, Claire explores the town's main shopping street and buys a Dracula cape for her stepdaughter Josefien. Back in her hotel she is interviewed by a female

journalist. I present a provisional, personal translation of a quote from chapter 6 of Part 1 'Claire' (cf. Van Dalen-Oskam [2021] 171).

The journalist fiddles with coffee, sugar, milk. She has remarkably graceful hands. At the same time, in a conversational tone of voice, she asks questions intended to break the ice. Crazy weather, isn't it, Mrs Pacman? A heat wave like this is very rare here. Is this her first visit to the town, has she had time to look around? The old town is certainly worth a look. As is the river, with on the bank the inn where Bram Stoker once stayed. And, of course, the bench on the cliff where Stoker, looking out to sea, got the idea to land his Count Dracula right here.

She nods, suppressing her boredom. It's all in her guidebook. It's strange how proud people in this seaside resort are that a fictitious monster washed ashore right there. But at least it inspires some snide remarks in the guise of literary allusions for her acceptance speech.

She takes a sip of her coffee, which is of moderate quality. You'll see that at least one photographer comes up with the idea of having her pose on Stoker's bench today. Well, she can wear Josefien's cape for the photo. And at this thought, out of the blue, she is overwhelmed by nostalgia for the time when she was casually rehearsing her stepdaughter's French homework at the kitchen table. Well, Claire, I'll call you Moon from now on. Followed by a screeching fit of laughter, with a lot of swishing of yard-long blonde hair.

The place is never mentioned by name but it is clear Claire is in Whitby. During her stay Dracula will return in another form. In between acts, we learn about the origin of Claire's marital problems, which turn out to revolve around a sensitive issue her stepdaughter brought up. In the second part of the novel we are presented with Axel's perspective and in the final part with that of Josefien, Axel's daughter and Claire's stepdaughter.

I quote the motivations of a couple of readers, again all women. A maximum score of 7 for literary quality and a 6 for general quality was motivated as follows: 'I find Renate Dorrestein's imagination highly original. Her writing style is concise, fast, surprising. She plays with language. Language seems to be an independent factor in her writing' (ID_5426). The next reader was less enthusiastic. She gave the book a 4 for literary quality and a 5 for general quality: 'Nothing more than a nice, smoothly written story' (ID_3869). The novel received a rating of 5 for both general and literary quality from the following reader: 'While the main character is staying abroad (and I think has just been abandoned by her husband or vice versa) she immerses herself in a river of thoughts and feelings. More is going on than appears at first glance and her personality also proves to be different from what the reader initially thinks' (ID_11525).

Renate Dorrestein not only wrote novels but also published her thoughts and experiences about writing, authorship, and being a female author in the Dutch literary landscape. In 2017, a big book titled *Onder de motorkap van het schrijverschap* (**Under the Bonnet of Authorship*) brought together two earlier texts: *Het geheim van de schrijver* (**The Writer's Secret*) from 2000 and *De blokkade* (**Writer's Block*), from 2013. The first of these aimed to be a handbook for aspiring writers, a guide for readers to understand their own reading taste, and a kind of literary autobiography (2017: 12). The second described how Dorrestein suffered and overcame a writer's block, which started right after *De stiefmoeder* (**The Stepmother*) was published to great acclaim. The high praise she received for that novel was a big surprise to her and she could not explain why this novel finally brought her success (2017: 234, 280–281). This may also indicate that the genesis of the novel did not differ from that of earlier ones. Dorrestein described herself in *Het geheim van de schrijver* as a so-called 'process writer', someone who starts with an initially small idea. During the creative process of writing and rewriting that follows she sees which turns the story can take and finds out whether the developing manuscript has the potential to grow into a novel. She believed that the story is in fact already there, and that it must find her as a writer – and not that she was the one who consciously designed and structured the story. She experienced the search for the story that already exists somewhere as an intensive and exciting process, with an uncertain outcome. For some of her novels, it took ages before she was sure it would work. Sometimes she had to bin a concept, and other times it took dozens of versions before it finally became a good novel.

But what makes a good and literary novel is not clear. Dorrestein was very much aware of the fact that it can take a while before a writer is read with pleasure and appreciation and that it is often unclear what exactly causes such a breakthrough. Has the work become better, or have the prevailing conventions changed? Dorrestein herself always explicitly stated that above all she wanted to tell a good story. She also often referred to the fact that for a long time it was actually a disqualification for a (female) author to be called a 'storyteller', but that recently, narrative prose gained considerable prestige (2017: 172). That is to say: *literary* prestige. Dorrestein emphasized that it is clear that many Dutch readers aspiring to be writers find literary quality important, but that it is not clear how exactly literary value is attributed (2017: 30–31). She summarized her view on how literary critics deal with the problem of literary quality as follows:

The choice of subject does play a role. Some consider motherhood to be a completely acceptable literary theme; others believe that we may only speak of real literature when, following Hemingway, people are out on the streets at night in a novel, throwing up in gutters and visiting prostitutes. And apart from that, a

novel, because of its emotional charge, makes both friends and enemies, just as an outspoken personality does.

There are literary theories that can be used to identify quality. However, there are so many different methods of exegesis in circulation that they too have become subjective, just like taste. Literature is not physics and cannot be captured in rules or formulas. A good novel, by definition, has an intangible element of *je ne sais quoi* in it, a certain organic magic that one is unable to point out. Even if one takes a magnifying glass to the page, a novel owes its brilliance to what is not there, to the text beneath the text' (2017: 31–32, my translation cf. Van Dalen-Oskam [2021] 173).

The role of topics or subjects of novels will be addressed in Chapter 6 of this monograph.

Participants in The National Reader Survey agreed with contemporary critics: *De stiefmoeder* scored more than 0.5 point (on a scale of 7) higher for literary quality than *De leesclub*. Measurements of the vocabulary and word frequencies showed that in this respect, the two novels are fundamentally different. This suggests a big difference in style, which may be related to the label of anti-novel Dorrestein herself later selected for this novel and for none of her other works. Dorrestein agreed that the style used to shape a story is very important and that style is a very personal thing.

All authors have an individual penmanship. The handwriting can be angular or flamboyant, baroque, or sparse. The stylistic signature of some is repetition, of others a mixture of elevated and banal terminology, and of others irony, but all writers are in a sense as stuck with their style as they are with the colour of their eyes. Authors do not have a choice in the matter, they have to make do with what they have been given by gods or genes. There is some polishing to be done, but basically style is as uncontrollable a phenomenon as left- or right-handedness.

Of all personal shortcomings, stylistic ones are the most difficult to stamp out. For years, I have been fighting with all my might against the phenomenon that on paper I often come across as jollier than intended (...). And every day I have to rein in the abundance of adjectives and adverbs that come at me like wasps to treacle' (2017: 139–140 cf. Van Dalen-Oskam [2021] 169–170).

A very specific stylistic element Dorrestein referred to is the grammatical time in which a novel is set. Dorrestein pointed out that in literary fiction, for a long time the past tense was favoured, but that nowadays most novels are written in the present tense.

It gives the story a semblance of topicality and urgency. The reader is on top of it. I suspect that the present tense has become so popular because it sounds more 'real' than the somewhat fairy-tale past tense ('Once upon a time there was a handsome prince'). Its use fits in seamlessly with the current preference for 'true stories' (2017: 148, cf. Van Dalen–Oskam [2021] 170).

She herself preferred the past tense, she wrote, but she explicitly stated she did not rule out that one day, she would go 'with the times' (2017: 149). Dorrestein wrote this in 2000, so it is not surprising that we do not read any comments about the two novels in our research corpus. It is interesting, however, to note that *De leesclub* (2010) uses past tense, whereas *De stiefmoeder* (2011) uses present tense. This indeed may be another reason why the two novels end up in different clusters in the measurements. Another feature, mentioned earlier, can also play a role: the first-person plural narrative of *De leesclub* ('we') and the third-person narrative of *De stiefmoeder*. To find out to which extent these two characteristics are responsible for the striking difference between Dorrestein's two novels requires larger-scale research with many more novels in the corpus than are currently available.

Dutch Originals Opposed

For a comparison between the Dutch Literary novels written by female authors that received the highest and lowest scores for literary quality I used the keyword function in AntConc. I compared the vocabulary of the six novels with the highest scores with the six lowest-scoring ones, leaving out the middle group of seven novels. Dorrestein's *De stiefmoeder* (**The Stepmother*) was part of the group with the highest scores. *De leesclub* (**The Reading Group*), in the middle group, was not included in this measurement.

In the six highest scoring novels *hij* (he) occurred more often than statistically probable, and the books with the lowest scores more often had *ik, me/mij, mijn, mezelf, jij, je, jou, jouw, and we* (I, me, my, myself, you, your, we). The novels with the highest literary quality contained relatively more verb forms in the past tense, such as *had, hadden, was, waren, kon, zag, keek, bleef, and wist* (had, had, was, were, could, saw, looked, stayed, knew) as well as two in the present tense, *zegt and denk* (says, thinks). The least literary novels had a higher frequency of many verb forms in the present tense compared to the highly literary novels: *heb, ben, is, word, wordt, wil, wilt, vraag, zeg, vind, neem, doe, leg, voel, reageer, sta, loop, and zie* (have, am, is, shall, will, want, want, ask, say, find, take, do, lay, feel, react, stand, walk, see). These tendencies may indicate that the literary high-ranking novels in this selection more often dealt with male characters (in a third-person narrative)

and the story was told in the past tense, while the low-ranking books more often featured a first-person narrative and/or contained more dialogue and were written in the present tense. Note that from this perspective, Dorrestein's *De stiefmoeder* seems to be an exception here.

Conclusion

In this chapter, I focused on the novels in the category Literary novel, so as to exclude differences in prestige of genres from the equation. And because novels written by female authors received significantly lower scores for literary quality than those written by men, I first had a closer look at the 34 translations and then at the 23 Dutch originals written by women. I presented a variety of readers' motivations for their ratings of the translated novel that received the highest mean score for literary quality, *Het grote huis* (*Great House*) by Nicole Krauss, as well as about one of the novels with the lowest ratings, Elizabeth Gilbert's *Eten, bidden, beminnen* (*Eat, Pray, Love*). I then compared the translations by means of several text analysis experiments. There, Emma Donoghue's novel *Kamer* (*Room*) was one of the outliers, and I hypothesized that the style of the novel, which carefully reflects the language of the five-year-old boy who is the main protagonist, was probably responsible for this.

The Dutch originals were discussed in somewhat more detail than the translated ones. I paid special attention to Margriet de Moor's high-scoring *De schilder en het meisje* (**The Painter and the Girl*), comparing both ratings and motivations of readers of this novel with those of the lowly rated *De mannestester* (**The Tester of Men*) by Heleen van Royen. The text analysis experiments done on the Dutch originals also yielded some interesting observations. I discussed the outlier *Vinexvrouwen* (**Middle-Class Women in Modern Suburbia*) and explored why the novels by mother and daughter Margriet de Moor and Marente de Moor clustered together. This led to the conclusion that different text analysis methods yielded slightly different results, and that although the novels by mother and daughter were very similar in their most frequent words, each of them was even more similar to the novel of an author of around the same age. Next, I tried to find out why the two novels by Renate Dorrestein, *De leesclub* (**The Reading Group*) and *De stiefmoeder* (**The Stepmother*), ended up far apart from each other, contextualizing Dorrestein's work as an author and as a coach in creative writing. The differences between Dorrestein's two novels may partly reflect the fact that one of them was characterized as an anti-novel by Dorrestein herself. Furthermore, a difference in tense may have played a role.

For both the translations and the Dutch originals I analysed the keywords, words that were statistically overrepresented in the highly valued novels compared to

the least valued book and vice versa. The translated novels with the highest scores for literary quality contained more pronouns suggesting a first-person narrative than those at the bottom of the ranked list. The group with low literary quality yielded a higher amount of second-person pronouns, which may be related to more dialogue. Furthermore, the tenses of verb forms in the translations showed the exact opposite pattern from that in the Dutch originals: the present tense is found more often in the most literary translated novels and the past tense more in the least literary ones. In Chapter 6 and 7 I will go into this observation in more detail. But first, in Chapter 5, I will report on the same analyses of the Literary novels written by male authors.

References

All hyperlinks were last accessed in February 2023.

- Anthony, Laurence. *AntConc*. Version 3.5.8., Tokyo: Waseda University, 2019, <https://www.laurenceanthony.net/software.html>.
- Dorrestein, Renate. 'Ouverture.' *De zondagmiddagauto* [*The Sunday Afternoon Car*, not available in English], B for Books Publishing, 2012, pp. 7–26.
- Dorrestein, Renate. *Onder de motorkap van het schrijverschap* [*Under the Bonnet of Authorship*, not available in English]. Em. Querido's Uitgeverijen, 2017.
- Eder, Maciej, et al. 'Stylometry with R: A Package for Computational Text Analysis.' *The R Journal*, vol. 8, no. 1, 2016, pp. 107–121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.
- 'Letterenfonds Vertalingendatabase' ['Dutch Foundation for Literature Translation Database']. *Dutch Foundation for Literature*, <https://www.letterenfonds.nl/nl/vertalingendatabase>.
- McCarthy, Rachel and James O'Sullivan. 'Who wrote Wuthering Heights?' *Digital Scholarship in the Humanities*, vol. 36, no. 2, 2021, pp. 383–391, doi:10.1093/llc/fqaa031.
- Rybicki, Jan. 'The Great Mystery of the (Almost) Invisible Translator: Stylometry in Translation.' *Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research*, edited by Michael P. Oakes en Meng Ji, John Benjamins Publishing Company, 2012, pp. 231–248.
- Van Dalen-Oskam, Karina. 'Names in Novels: An Experiment in Computational Stylistics.' *LLC: The Journal of Digital Scholarship in the Humanities*, vol. 28, no. 2, 2013, pp. 359–370, doi: 10.1093/llc/fqs007.
- Van Dalen-Oskam, Karina. *Het raadsel literatuur: Is literaire kwaliteit meetbaar?* [*The Riddle of Literature: Can We Measure Literary Quality?*] Amsterdam University Press, 2021, <https://karinavdo.github.io/RaadselLiteratuur/>.

5. Literary Novels Written by Men

From Barnes to Grunberg

Abstract: Literary novels written by male authors received higher scores for literary quality than those by female authors. This chapter focuses on the literary novels written by men, both translations and Dutch originals. Cluster analysis using Burrows' Delta, principal components analysis (PCA), and keyword analysis are applied to explore differences in word usage and word frequencies between highly rated and lowly rated novels to search for an explanation of the ratings. The translation of Julian Barnes' *The Sense of an Ending*, which received the highest mean score for literary quality, gets special attention. Dutch novels analysed in more detail are by Mart Smeets, Herman Koch, Kluun, and Arnon Grunberg, whose novels linguistically were close to Romance novels but received much higher ratings.

Keywords: Literary quality, Julian Barnes, Herman Koch, Arnon Grunberg, Computational Literary Studies, Distant Reading

Barnes and the Booker

This chapter follows roughly the same pattern as the previous one, but instead of Literary novels written by female authors it features those written by men. They are kept apart because the ratings of the participants in The National Reader Survey showed a certain bias against Literary novels authored by women, as I showed in Chapter 3 of this book. This suggested that author gender may be a confounding factor when we want to find trends in perceptions of literariness, and that the corpus should be adapted accordingly. I will first show where, in relation to the other main categories of Suspense and Romance, the male authored Literary novels ended up in the overall ranking. And since the knowledge that a book is a translation may also have played a role in readers' assessment of literary quality, I will first analyse the translations and then the novels written in Dutch. I will bring the Dutch Literary novels by women and men together again in Chapter 6 for a final experiment.

Of the 401 novels in the research corpus, 192 were written by male authors. Mixed duos were excluded from the discussion in this chapter. Of these 192, 90 were labelled as Literary novel, 91 as Suspense, and 11 as Other. The Romance category did not feature any books solely written by men. In the literary ranking, Literary novels by male writers dominated ranks 1–127. Male-authored Suspense mainly occurred between ranks 128–386.

A small group of five Suspense novels ended up between the Literary novels in the top 128: C.J. Sansom, *Winter in Madrid (Winter in Madrid)* at rank 89, John le Carré's *Ons soort verrader (Our Kind of Traitor)* at 96, and three Suspense novels by Henning Mankell *Kennedy's brein (Kennedy's Brain)*, *De gekwelde man (The Troubled Man)*, and *De Chinees (The Man from Beijing)* at rank 112, 119 and 121, respectively. Mankell was also represented with two Literary novels; they received slightly higher scores for literary quality than his Suspense novels and are positioned at rank 98 (*De Daisy Sister – *Daisy Sisters*; the asterisk indicates no English translation is available) and 105 (*De geschiedenis van een gevallen engel – A Treacherous Paradise*). In Chapter 3 I have explored Mankell's novels across genre borders in more detail.

Literary novels written by male authors also occur in the part of the literary ranking that is dominated by Suspense. There are nine of these, the first at rank 132 and the last at 361:

- Ken Follett, *Val der titanen (Fall of Giants)*
- Robert Vuijsje, *Alleen maar nette mensen (*Decent People Only)*
- Rik Launspach, *1953 (*1953)*
- Justin Cronin, *De oversteek (The Passage)*
- Nicholas Evans, *De vergeving (The Brave)*
- David Baldacci, *Die zomer (One Summer)*
- James Worthy, *James Worthy (*James Worthy)*
- Kluun, *Haantjes (*Alpha Males)*
- Mart Smeets, *De afrekening (*The Reckoning)*

Literary novels originally written in Dutch are spread throughout the literary quality ranks. It is noteworthy that the three highest ranked Dutch-language novels are written by Flemish authors, so by authors originating from Flanders, part of the Netherlands' neighbouring country in the South: Belgium.

The forty translated Literary novels came from eight different languages: almost half, nineteen, from English, six from Italian, four from Spanish, three from Swedish, and two from German, French, Japanese, and Portuguese. English is found throughout the ranked list. French and Japanese appear only at the top of the literary ranking, and Swedish only in the least valued group. Readers who

would like to explore the ranked list in more detail are referred to Appendix 2 and to the litRiddle package for R, see Appendix 3.

The highest score for literary quality was reserved for *Alsof het voorbij is* (*The Sense of an Ending*) by Julian Barnes. Rank 2 up until 31 were also taken by Literary novels authored by male writers. *Alsof het voorbij is* was published in Dutch in October 2011 and was translated by Ronald Vlek. The English original *The Sense of an Ending* also appeared in 2011, and in October of the same year received the prestigious Man Booker Prize for the best novel written in English and published in the United Kingdom or Ireland. The Dutch translation of the novel became available in the same month the prize was awarded, and it is very likely that Dutch readers were aware of the award and – consciously or unconsciously – took it into account in their assessment of the novel's literary quality. The 817 received ratings resulted in an average score of 6.62 for both literary quality and general quality, where 7 was the highest score a reader could give to a book. In this, the participants showed a lot of agreement, as can be seen in the visualization of the distribution of the ratings in Figure 5.1.

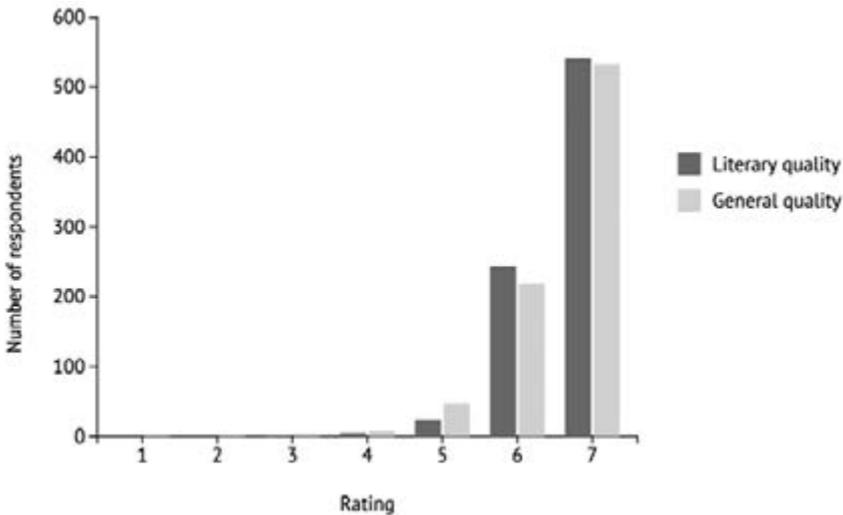


Figure 5.1: Ratings of Julian Barnes, *Alsof het voorbij is* (*The Sense of an Ending*)

Alsof het voorbij is (*The Sense of an Ending*) is narrated by the elderly Tony Webster. He describes the childhood friendship with his classmates Colin and Alex and later the gifted Adrian. But after secondary school they begin to drift apart. Adrian gets involved in a relationship with Veronica Ford, who used to be Tony's girlfriend. Then the friends are informed that Adrian has committed suicide. Years later, Tony

receives word that he has inherited a small sum of money and some documents from Veronica's recently deceased mother. This reactivates Tony's memory in an unexpected way; he must thoroughly revise his judgement of himself and others. One of the participants wrote: 'This is the best book I have read in ages. It is very beautifully written. The subject matter is very surprising and very well developed. The book stays in your mind for a very long time' (ID_8819).

A total of 114 participants motivated their rating of the literary quality of the novel. Per usual, the writing style and the 'deeper layers' of the novel are frequently attributed as literary properties. This reader rated the novel with a 6 for literary quality: 'The writing style and the fact that you can talk about it in a reading group' (ID_5788) – the previous chapter referred to a similar motivation for the highest-ranking Literary novel by a female author, *Het grote huis* (*Great House*) by Nicole Krauss. The next reader also gave Barnes' novel a 6 for literary quality, and was one of the few who combined a high score with a negative comment: 'Expertly written fiction that, in this case, I actually don't like' (ID_1640); as for the other substantiations by respondents I quote the complete text this reader put in, so we do not get to know why the respondent did not like the novel. The vast majority of the readers who motivated their score of the book gave it the top score 7: 'I found it a very intense book of which I enjoyed every sentence and that left a deep impression' (ID_6627). 'It is beautifully described how the main character finds out that his memories are faltering' (ID_13162). 'It is beautifully written, has a surprising storyline and when I finished it, I started to read it again' (ID_9309).

In the next sections of this chapter, a variety of other novels from this subcategory will be discussed in more detail, starting with translations.

Translations Opposed

To find out how the forty translated Literary novels by male authors compare to each other in their vocabulary and word frequencies, I divided them into three groups of roughly the same size: those with the highest mean scores for literary quality, those with the lowest scores, and the group in between. The borders between the three groups are arbitrary, because in contrast to the 34 translated Literary novels by female authors, high scores for literary quality predominate.

Group H: highest scores for literary quality – mean score

Julian Barnes, *Alsof het voorbij is* (*The Sense of an Ending*) – 6.62

Michel Houellebecq, *De kaart en het gebied* (*The Map and the Territory*) – 6.41

Haruki Murakami, *Norwegian Wood* (*Norwegian Wood*) – 6.31

Umberto Eco, *De begraafplaats van Praag* (*The Prague Cemetery*) – 6.24

- David Mitchell, *De niet verhoorde gebeden van Jacob de Zoet* (*The Thousand Autumns of Jacob de Zoet*) – 6.19
 Haruki Murakami, *1Q84* (*1Q84*) – 6.19
 Jonathan Franzen, *De correcties* (*The Corrections*) – 6.16
 Jonathan Franzen, *Vrijheid* (*Freedom*) – 6.15
 Laurent Binet, *HhhH* (*HHhH*) – 6.02
 John Irving, *In een mens* (*In One Person*) – 5.88
 Chad Harbach, *De kunst van het veldspel* (*The Art of Fielding*) – 5.88
 Paolo Giordano, *De eenzaamheid van de priemgetallen* (*The Solitude of Prime Numbers*) – 5.88
 Charles Lewinsky, *De verborgen geschiedenis van Courtillon* (**Saint John's Eve*) – 5.86

Group M: middle scores for literary quality – mean score

- Carlos Ruiz Zafón, *De gevangene van de hemel* (*The Prisoner of Heaven*) – 5.80
 John Irving, *De laatste nacht in Twisted River* (*Last Night in Twisted River*) – 5.77
 Charles Lewinsky, *Het lot van de familie Meijer* (*Melnitz*) – 5.72
 Carlos Ruiz Zafón, *Het spel van de engel* (*The Angel's Game*) – 5.66
 Dave Eggers, *Wat is de wat* (*What is the What*) – 5.64
 Irvin D. Yalom, *Het raadsel Spinoza* (*The Spinoza Problem*) – 5.63
 Niccolò Ammaniti, *Ik en jij* (*Me and You*) – 5.61
 Niccolò Ammaniti, *Zo God het wil* (*As God Commands*) – 5.56
 Niccolò Ammaniti, *Het laatste oudejaar van de mensheid* (**The Last New Year's Eve of Humanity*) – 5.49
 Paulo Coelho, *De beschermengel* (*The Valkyries*) – 5.48
 Carlos Ruiz Zafón, *De schaduw van de wind* (*The Shadow of the Wind*) – 5.46
 Karl Marlantes, *Matterhorn* (*Matterhorn*) – 5.42
 Khaled Hosseini, *Duizend schitterende zonnen* (*A Thousand Splendid Suns*) – 5.41
 Chris Cleave, *Kleine Bij* (*The Other Hand*) – 5.36

Group L: lowest scores for literary quality – mean score

- Niccolò Ammaniti, *Laat het feest beginnen* (*Let the Games Begin*) – 5.33
 Lawrence Hill, *Het negerboek* (*The Book of Negroes*) – 5.32
 Paulo Coelho, *Aleph* (*Aleph*) – 5.30
 John Boyne, *Het winterpaleis* (*The House Of Special Purpose*) – 5.27
 Ildefonso Falcones, *De hand van Fatima* (*The Hand of Fatima*) – 5.17
 Henning Mankell, *De Daisy Sisters* (**Daisy Sisters*) – 5.00
 Henning Mankell, *De geschiedenis van een gevallen engel* (*A Treacherous Paradise*) – 4.92
 Jonas Jonasson, *De 100-jarige man die uit het raam klom en verdween* (*The 100-Year-Old Man Who Climbed Out the Window and Disappeared*) – 4.81

- George Mastras, *Tranen over Kashmir (Fidali's Way)* – 4.79
 Ken Follett, *Val der titanen (Fall of Giants)* – 4.57
 Justin Cronin, *De oversteek (The Passage)* – 4.18
 Nicholas Evans, *De vergeving (The Brave)* – 4.16
 David Baldacci, *Die zomer (One Summer)* – 4.09

As in Chapter 4 I started the text analysis experiments with several principal components analyses (PCA). I made use of the Stylo package for R (Eder et al.). Figure 5.2 shows a visualization of the results of a PCA based on the thousand most frequent words in the texts. Each group has its own grey scale, to help to easily spot any major differences. More about the method can be found in Chapter 3. The essence is that the closer books are placed to each other in the visualization, the more similar they are in the frequency distribution of their one thousand most frequent words. As for women's Literary novels in translation and in Dutch, the three groups with the highest and lowest literary quality scores and the group in between show an uneven spread over the entire field of the graph. This means that here too, the difference between the most and least literary translations seemingly cannot be related to the use of the thousand most frequent words. There must be more to it than word choice and word frequencies alone.

An outlier at the bottom of the graph is *De hand van Fatima (The Hand of Fatima)* by Ildefonso Falcones from the group with the lowest scores for literary quality. As of yet, it is unclear why this novel ended up in this isolated place. Falcones' work has an exceptional size of 944 densely printed pages, but whereas it is commonly accepted as understandable that very short texts end up as outliers, the same is not true for very long ones. Still, when I redid the analysis with random samples of different lengths from all forty novels, the book was no longer an outlier (graphs not shown).

Outliers at the top are *Norwegian Wood (Norwegian Wood)* by Haruki Murakami from the group considered most literary, and *De Daisy Sisters (*Daisy Sisters)* by Henning Mankell from the least literary group. In Chapter 3 I paid special attention to the five novels by Henning Mankell in the corpus, and showed how *De Daisy Sisters (*Daisy Sisters)*, one of his two Literary novels, deviated from the other Literary novels and the three Suspense books in a similar way. I hypothesized that this may be caused by the fact that this Dutch translation from 2009 goes back to a significantly older Swedish original from 1982. The same is true for Haruki Murakami's *Norwegian Wood (Norwegian Wood)*, which was originally published in Japanese in 1987 and was only published in a Dutch translation in 2009. The distance between the two novels by Murakami is much smaller than that between the two by Mankell. Whether the original publication date of the two novels is responsible for their outlier positions cannot

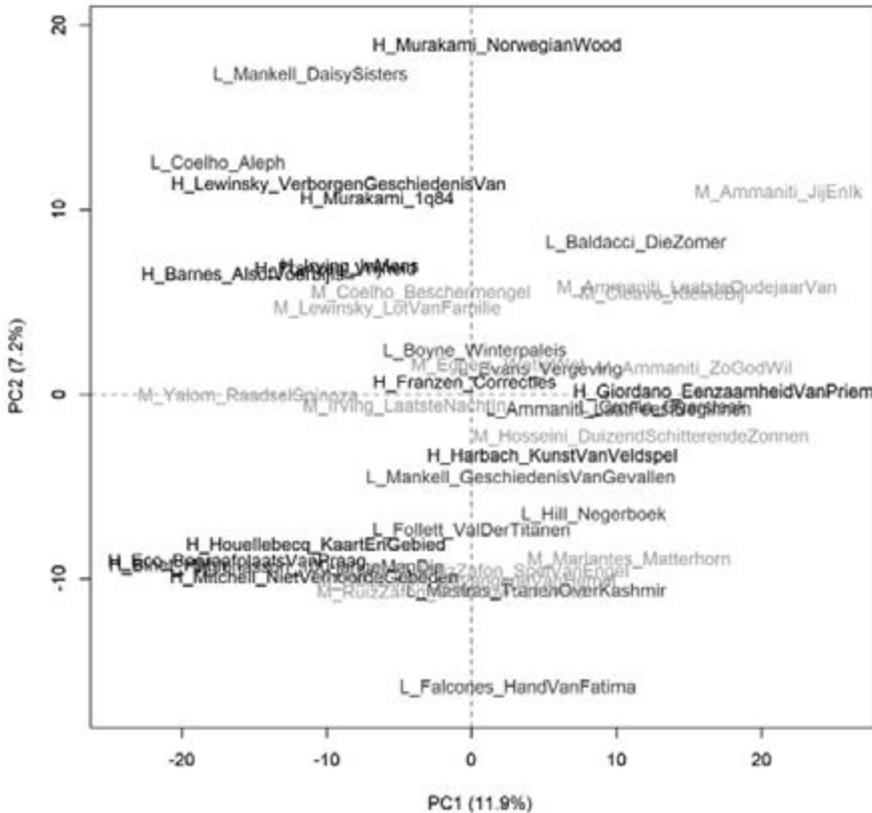


Figure 5.2: Translated Literary novels written by male authors, PCA (1000 most frequent words). Scores for literary quality: H_ (highest), L_ (lowest), and M_ (middle group).

be said with certainty, but to me it does seem probable. More tests are needed, however, using a much larger corpus with representatives from different time periods.

Next, I used the keyword function in Laurence Anthony's AntConc tool to find out which words occurred significantly more in the novels with the highest scores for literary quality than those with the lowest scores, and vice versa. I excluded the middle group from this comparison. In the novels with the highest literary quality we see a relatively greater use of the personal pronouns *ik* (I), *je* (you), and *mij* (me), and of the present tense in the verb forms *is* (is) and *heb* (have). In the novels that were rated least literary, other personal pronouns were used more prominently: *ze*, which can be either a feminine singular (she, her) or a plural (they, them), *hen* (them), *hij* (he), and *hem* (him). The lowest-rated group featured some verb forms in the past tense much more frequently than expected: *keek* (looked), *zag* (saw),

antwoordde (replied), *riep* (yelled), and *waren* (were). Novels with the highest scores for literary quality seem to tend to be first-person narration due to their relatively larger share of *ik* (I) and *mij* (me). Another explanation may be that they have more dialogue because of the more prominent occurrence of *je* (you). The books that received the lowest scores for literary quality more often seem to be written in the past tense and as a third-person narrative.

In the novels with the highest literary quality, the following words are used much more frequently than in the novels with the lowest literary quality: *doch* (yet), *als* (if), *ook* (also), *of* (or), and *dus* (so). Most of these are conjunctions. More prominent in the least literary group are *de* (the) and *en* (and). As such, there seems to be a significant difference in the use of conjunctions in the highest and lowest literary quality categories. The conjunction *en* (and) is more prominent in the novels with least literary quality, as opposed to *of* (or) and *doch* (yet) in the novels scoring highest for literary quality.

The Dutch conjunction *doch* (yet) is archaic and therefore invites further inspection.¹ A closer look showed that it only occurs in nine of the forty novels, with a total of 323 occurrences. In eight of these it is only used once or twice, and the remaining 304 all come from David Mitchell's historical novel *De niet verhoorde gebeden van Jacob de Zoet* (*The Thousand Autumns of Jacob de Zoet*). The novel about a Dutch bookkeeper on Deshima is set in Japan in the period 1799 to 1811, and was translated into Dutch by Harm Damsma and Niek Miedema. Below, I present the first two occurrences of *doch* (on the first two pages of the novel) in both translation and English original.

Door het mousseline gordijn heen fluistert de geneesheer Maeno: 'Ik wilde mij er zelf van vergewissen hoe het kind ligt, **doch**' – de bejaarde geleerde kiest zijn woorden met zorg – 'dit schijnt niet gepermitteerd.'

Dr Maeno whispers through the muslin curtain. 'I wanted to examine the child's presentation myself, but...' the elderly scholar chooses his words with care '... but this is prohibited, it seems.'

'Dokter Maeno, bent u het met me eens dat hier sprake is' – ze gebruikt het Hollandse woord – 'van een "dwarsligging"?'

'Wellicht,' zegt de geneesheer in hun geheimtaal, '**doch** zonder nader onderzoek ...'
'Dr Maeno, would you agree the infant is in' – she uses the Dutch term – 'the "transverse breech position"?'

'Maybe,' the doctor answers in their code-tongue, 'but without an examination...'

1 *Doch* is one of the words on the list of archaic words of the Genootschap Onze Taal, the Dutch language society, see <https://onzetaal.nl/taaloket/modern-taalgebruik>.

So it seems that the translators chose the somewhat archaic Dutch *doch* to represent the English non-archaic 'but'. I suspect that Damsma and Miedema did this to compensate for comparable English archaisms elsewhere that could not be properly translated into equally archaic Dutch. In this way, the overall style of the translation still reflects that of the original.

Dutch Originals Compared

The corpus consisted of fifty male-authored Literary novels written in Dutch. Female authors had 23 novels in the list. Men therefore did considerably better on the Dutch bestseller list and in the ranked list of loans in Dutch public libraries that informed the selection of novels to be included in the corpus. And what is more: where at the time of writing of this book only one of the 23 Literary novels by women was translated into English, twelve novels by male authors were available in an English translation. This is a significant difference. As before, I have marked the titles I translated into English myself with an asterisk, as can be seen in the overview of the three groups below.

Group H: highest scores for literary quality – mean score

- Erwin Mortier, *Godenslaap* (*While the Gods Were Sleeping*) – 6.60
- Erwin Mortier, *Gestameld liedboek* (*Stammered Songbook*) – 6.54
- Tom Lanoye, *Sprakeloos* (*Speechless*) – 6.37
- A.F.Th. van der Heijden, *Tonio* (**Tonio*) – 6.29
- Stephan Enter, *Grip* (**Grip*) – 6.26
- J. Bernlef, *Geleende levens* (**Borrowed Lives*) – 6.25
- Tommy Wieringa, *Caesarion* (*Caesarion*) – 6.20
- Gerbrand Bakker, *De omweg* (*The Detour*) – 6.19
- Arthur Japin, *De overgave* (**Surrender*) – 6.17
- Thomas Rosenboom, *Zoete mond* (**A Mouth Full of Sweets*) – 6.17
- Thomas Rosenboom, *Mechanica*† (**Mechanics*) – 6.16
- Arnon Grunberg, *Huid en haar* (**Hair and Hide*) – 6.16
- J. Bernlef, *De een zijn dood* (**One Man's Meat*) – 6.14
- Arnon Grunberg, *Selmonosky's droom*† (**Selmonosky's Dream*) – 6.13
- Arthur Japin, *Vaslav* (**Vaslav*) – 6.11
- Arnon Grunberg, *De man zonder ziekte* (**The Man Without Illness*) – 6.11
- Stefan Brijs, *Post voor mevrouw Bromley* (**Mail for Miss Bromley*) – 6.09

Group M: middle scores for literary quality – mean score

- P.F. Thomése, *De weldoener* (**The Benefactor*) – 6.09
- Adriaan van Dis, *Tikkop* (*Betrayal*) – 6.07

- J.J. Voskuil, *De buurman* (**The Neighbour*) – 6.05
 Arthur Japin, *Dooit* (**Thaw*) – 6.03
 Bernard Dewulf, *Kleine dagen* (**Little Days*) – 5.98
 F. Springer, *Quadriga* (**Quadriga*) – 5.95
 Jan Siebelink, *Het lichaam van Clara* (**Clara's Body*) – 5.94
 Maarten 't Hart, *Verlovingstijd* (**Betrothal Time*) – 5.86
 Peter Buwalda, *Bonita Avenue* (*Bonita Avenue*) – 5.84
 Kader Abdolah, *De koning* (*The King*) – 5.84
 Tom Lanoye, *Heldere hemel* (**Clear Sky*) – 5.83
 Dimitri Verhulst, *De laatste liefde van mijn moeder* (**The Last Love of My Mother*)
 – 5.82
 Jan Siebelink, *Oscar* (**Oscar*) – 5.81
 Kees van Beijnum, *Een soort familie* (**A Kind of Family*) – 5.76
 Leon de Winter, *Recht op terugkeer* (**The Right to Return*) – 5.75
 Kader Abdolah, *De kraai* (**The Crow*) – 5.51
 Joost Zwagerman, *Duel* (**Duel*) – 5.50

Group L: lowest scores for literary quality – mean score

- Leon de Winter, *VSV of daden van onbaatzuchtigheid* (**Acts of Selflessness*) – 5.47
 Ronald Giphart, *Ijsland* (**Iceland*) – 5.34
 Remco Campert, *Dagboek van een poes* (**Diary of a Cat*) – 5.33
 Herman Koch, *Het diner* (*The Dinner*) – 5.12
 Alex van Galen, *Süskind* (**Süskind*) – 5.10
 Herman Koch, *Zomerhuis met zwembad* (*Summerhouse with Swimming Pool*) – 5.09
 Jaap Scholten, *Kameraad Baron* (*Comrade Baron*) – 4.97
 Nico Dijkshoorn, *Nooit ziek geweest* (**Never Been Ill*) – 4.79
 Ravelli, *De Vliegenvanger* (**Fly Trap*) – 4.75
 Ernest van der Kwast, *Mama Tandoori* (*Mama Tandoori*) – 4.73
 Ronald Giphart, *Zeven jaar goede seks*† (**Seven Years of Good Sex*) – 4.65
 Robert Vuijsje, *Alleen maar nette mensen* (**Decent People Only*) – 4.46
 Rik Launspach, *1953* (**1953*) – 4.29
 James Worthy, *James Worthy* (**James Worthy*) – 3.79
 Kluun, *Haantjes* (**Alpha Males*) – 3.50
 Mart Smeets, *De afrekening* (**The Reckoning*) – 3.06

As for the Literary novels written in Dutch by female authors, the list contains four titles that consist of separately published stories. They are marked with a cross. I have not included them in the analyses I report on in this chapter, because their relative shortness would make them outliers, without providing any other useful information.

The Dutch-language Literary novel receiving the highest score for literary quality is *Godenslaap* (*While the Gods Were Sleeping*) by Erwin Mortier, presented as an old woman named Heleen's retrospective on her youth during the First World War. It ended in second place on the literary ranking, right after *Alsof het voorbij is* (*The Sense of an Ending*) by Julian Barnes. The third place was also for Erwin Mortier, with his *Gestameld liedboek* (*Stammered Songbook*). This autobiographical novel sketches how the 66-year-old mother of the author suffers from Alzheimer's disease. In short notes, ranging from one sentence or a couple of sentences to a few pages, Mortier records observations of his mother and her illness. He shows how more and more words disappear from her memory, and how her language ability fades away. The many blank lines on the pages visually echo the description of this process. The quote presented below from the English translation by Paul Vincent from around a quarter into the book shows a scene from the first stages of her disease (cf. Van Dalen-Oskam [2021] 199):

I'd rather not imagine, but I do anyway: her walking doglike after him, the only person she still recognizes and on whom she fixates like a dove on the dovecote. Him first putting her on the sofa, picking up the vacuum cleaner again, but then giving up after she comes trailing after him for the umpteenth time.

The macabre dance of the two of them behind the vacuum cleaner round the table, she getting her feet caught in the cable and falling to the floor. Her crying. The blood. His feeling of guilt. The doctor.

I call back and ask: how can you stand it, I'm all in after a quarter of an hour.

He says: It's not that bad, I can take it.'

Most of the 576 participants in The National Reader Survey who rated the novel gave it the maximum score of 7 for literary quality (Figure 5.3). We also received 105 substantiations of scores. A reader aged 28 wrote: 'In a very beautiful way, with special phrasing, Erwin Mortier tries to make clear how he felt and what he observed during his mother's Alzheimer's' (ID_12416). And a 73-year-old man stated: 'Self-mockery and irony wrapped in Belgian humour. Mortier writes beautiful sentences. Belgians are often better at this than the Dutch' (ID_4543). A man aged 58 pointed out: 'Very subtle use of language, jumps in time, now and then it may be too much, especially when compared to Lanoye's *Sprakeloos* [*Speechless*]' (ID_9659). And this sixty-year-old man mentions Lanoye as well: 'Because of Erwin Mortier's very poetic use of language. Mortier knows, almost like no other, how to unlock a world of meaning in sparse text. Wonderful literature. (Just like his compatriot Tom Lanoye)' (ID_4974).

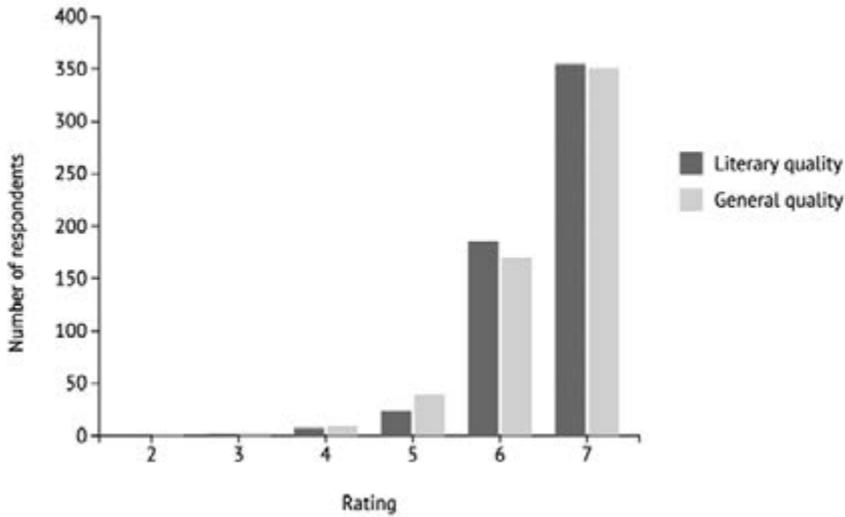


Figure 5.3: Ratings of Erwin Mortier, *Gestameld liedboek* (*Stammered Songbook*)

These motivations automatically lead us to a comparison with *Sprakeloos* (2009) by Tom Lanoye, which landed on the fifth place in the literary ranking. It was translated into English as *Speechless*, also by Paul Vincent. The subject is closely related to that of *Gestameld liedboek* (*Stammered Songbook*). Lanoye's flamboyant mother, a well-spoken and enterprising amateur actress nearing her eighties, suffers a stroke that seriously impairs her speech. Lanoye writes about the personality of mother Josée, the relationship with her husband Roger, and what their last years looked like. Lanoye's book is much longer than Mortier's. The chosen form is clearly different, although, as already mentioned, the subject is very similar. The following quote from *Sprakeloos* towards the end of the book may well illustrate the difference with *Gestameld liedboek*. The author visits his mother in the institution where she has been admitted when she can no longer be cared for at home (cf. Van Dalen-Oskam [2021] 201).

I enter her half-darkened room, prepared for the worst. All the same I am alarmed by the picture and the sound. She is sitting at the head end of the bed, with her back almost against the headboard and with a pillow on her lap that she is clinging to. She is rocking her body maniacally but softly backward and forward. It is her face that is most frightening. It is like a Greek mask, that of despair. She opens her mouth wide as can be and makes a quiet whining sound without crying and looks straight ahead, bewildered. With those deep furrows in her face, reinforced by the shadow of the curtains, and with her bulging eyes, her hollow cheeks and dishevelled hair she is reminiscent of a photo from the past: she as a judge, in

that avant-garde play *The Sentence*. But here there is no more acting involved. What she is embodying here, without trying to show or interpret, is the rock bottom of an existence, the end of the line. Perhaps she only now realizes what has befallen her.

For *Sprakeloos* (*Speechless*) I also present four motivations from readers. All awarded the book the highest score, 7, for literary quality. A 72-year-old reader justified his score in these words: 'Because the story is almost irrelevant, the use of language and sentence structure became reason enough for me to read on, such a book is powered by the writer's talent' (ID_10665). The following three substantiations were all by people in their twenties. 'Few things are explicitly explained, sometimes tiring to read because of sentence structure or word usage' (ID_1714). 'Skilfully plays with form, asks questions about literature through literature – a unique combination of form and content, transgressive and always beautiful' (ID_1823). And finally: 'Style, roaring Flemish sentences, theme of loss and transition' (ID_10913).

The Dutch Literary novel with the lowest score for literary quality was *De afrekening* (**The Reckoning*) by Mart Smeets, who for Dutch readers is a well-known former sports reporter. The readers' scoring pattern in Figure 5.4 looks completely different from what we have seen before for several highly rated novels and is even more extreme than what we observed in the previous chapter for *De mannestester* (**The Tester of Men*) by Heleen van Royen.

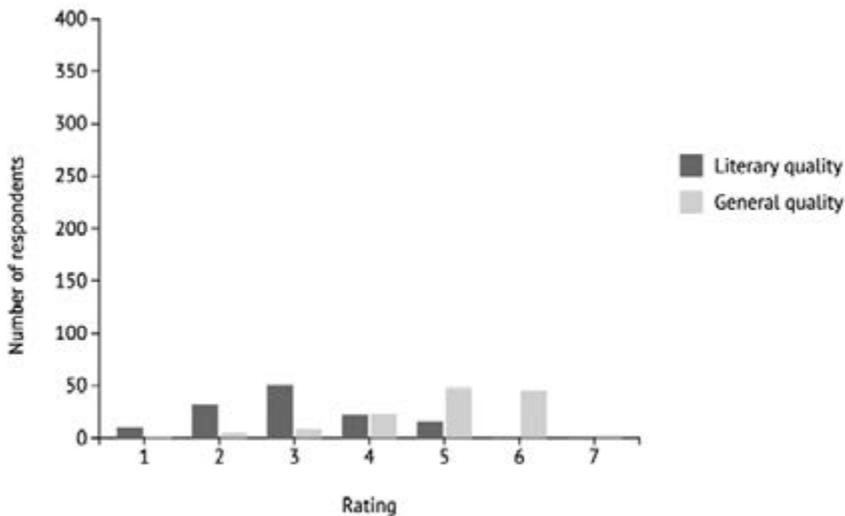


Figure 5.4: Ratings of Mart Smeets, *De afrekening* (**The Reckoning*)

De afrekening (**The Reckoning*) by Smeets is described on the cover and title page as a *wielroman* (cycling novel). The story is about sports reporter Stijn Miller, who has to deal with male cyclists, French organizers of the Tour de France, male and female Dutch celebrities, and all kinds of energetic and seductive women. The novel follows the life of Stijn in those years that the Tour de France started in the Netherlands. In the final episode, Stijn is a sixty-year-old with great professional prestige, but that year the Tour's hectic and social complications suddenly intensify to such an extent that he does not know how to handle this. In total, 129 participants in The National Reader Survey reviewed Smeets's novel and all their scores amounted to an average of 3.06 for *De afrekening's* literary quality. For general quality this was 4.96, considerably higher. Most of the 26 motivations were from men. One of the women wrote, substantiating her score of 3 for literary quality and 4 for general quality: 'I see it as non-fiction because I know the writer as a sports reporter, the rest then is very "artificial"' (ID_7174). This man gave the same scores, but motivated his judgement slightly differently: 'Well written, but no literary style and highly realistic' (ID_12020). Another man gave a 3 for general quality and a 2 for literary quality. Like the female reader just cited, he seems to have difficulty separating Smeets' well-known personality as a sports journalist from the author and narrator of the novel: 'It is a good and quickly told story, but journalistic rather than literary' (ID_5).

It seems that Smeets' fame in the Dutch world of sports journalism had a firm negative hold on the appreciation of his literary achievements. The fact that literature and sports do not go together has earlier been observed by researchers. Pieter Verstraeten studied the period around 1900 and observed that sport and literature reacted in opposite ways to processes of democratization. At the beginning of the nineteenth century, sports were primarily a pastime of the elite that gradually increased in popular appeal, whereas literature started out by reaching an increasingly wider public but over time targeted the artistic elite (Verstraeten 209). Sports and literary prestige are still not an accepted combination – although it does depend on what kind of sport: in Marente de Moor's highly praised literary novel *De Nederlandse maagd* (*The Dutch Maiden*) discussed in Chapter 4, the fencing sport is a perfect match with high literary prestige. For Mart Smeets' cycling novel the opposite is true.

To see how all Literary novels written by male Dutch authors compare to each other in their word use and word frequencies I again applied principal components analysis. As before, I made use of the Stylo package for R. Figure 5.5 presents the visualization of the measurement results of the thousand most frequent words. It displays the mixed distribution of the representatives of the three different groups (highest and lowest literary quality and the middle group). As always, one should remember that this only shows how the selected novels compare to each other. As soon as the composition of the selection changes, when novels are deleted, added, or replaced, the picture will shift accordingly because the overall bag of words

and the relative frequencies will also change. The visualization only shows the first two components, representing features that most explain the word frequency differences between the novels. The visualization helps to locate outliers and may help to find new ways of looking at the novels for a closer analysis using other tools.

In the centre of Figure 5.5, we again see that most of the novels from the groups with the highest and the lowest scores for literary quality, and from the group in between, merrily cluster together – literary quality as experienced by the readers cannot be read at a glance from differences in word usage and word frequencies. However, the top of the figure appears to be dominated by highly literary novels. This seems to be related to the Flemish background of a number of the authors involved. The two novels by Flemish author Mortier *Godenslaap* (*While the Gods Were Sleeping*) and *Gestameld liedboek* (*Stammered Songbook*) both appear at the top edge of the visualization, and not too far from these we find two novels by his compatriot Tom Lanoye – in addition to the already described *Sprakeloos* (*Speechless*) we have his 2012 book week gift *Heldere hemel* (**Clear Sky*). Another Fleming rubs shoulders with Mortier: Bernard Dewulf's *Kleine dagen* (**Little Days*). Of the authors from the Netherlands, it is P.F. Thomése whose use of words in *De weldoener* (**The Benefactor*) most resembles that of the Flemish authors; Kees van Beijnum's *Een soort familie* (**A Kind of Family*) also pops up here. At the bottom of the graph we see a preponderance of novels least appreciated for their literary quality. Apart from the already discussed *De afrekening* (**The Reckoning*) by Mart Smeets these include *Alleen maar nette mensen* (**Decent People Only*) by Robert Vuijsje, *Süskind* (**Süskind*) by Alex van Galen, and *Nooit ziek geweest* (**Never Been Ill*) by Nico Dijkshoorn. Kluun's *Haantjes* (**Alpha Males*) also appears in this vicinity. I will discuss this novel in more detail in the next section, combined with Herman Koch's *Zomerhuis met zwembad* (*Summerhouse with Swimming Pool*) which is hidden in the cluster in the middle of Figure 5.5.

The bottom of Figure 5.5 shows an isolated member of the middle group, *De buurman* (**The Neighbour*) by J.J. Voskuil. The novel is about Maarten and Nicolien Koning, known from Voskuil's earlier novels *Bijnader inzien* (**On a closer look*) from 1963 and the immensely popular seven-volume *Het Bureau* (**The Bureau*) from 1996–2000. In *De buurman* (**The Neighbour*), Maarten is the first-person narrator. Maarten and Nicolien have a new neighbour, a very reclusive man who makes sure he never runs into them. That is until his boyfriend moves in with him, who initiates more contact between the neighbours. I present a quote from the beginning of the novel in my own tentative translation. The novel as a whole contains an exceptionally large number of dialogues, such as the following (cf. Van Dalen-Oskam [2021] 208):

'Stallinga has a boyfriend,' I told her.
'How do you figure that?' she asked.

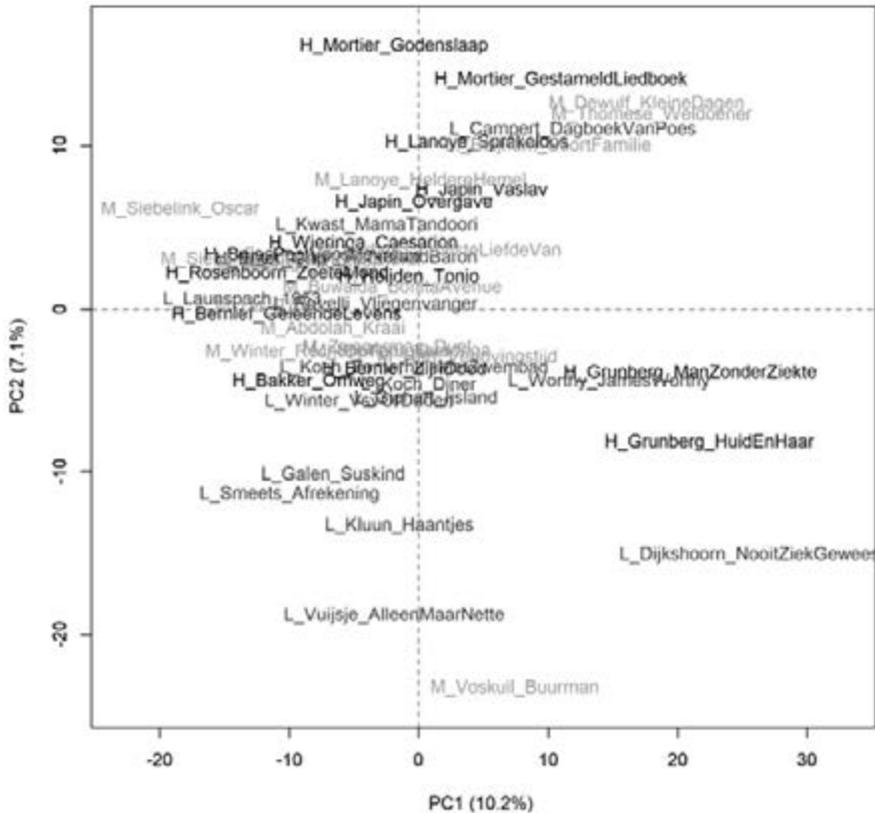


Figure 5.5: Dutch Literary novels written by male authors, PCA (1000 most frequent words). Scores for literary quality: H_ (highest), L_ (lowest), and M_ (middle group).

'A boy went upstairs with a mattress on his back.'
 'But that doesn't have to be a boyfriend, does it?'
 'According to Freek, it's his boyfriend.'
 'Yes, Freek! Of course he likes to say that. He thinks everyone is gay. That's what he wants.'
 'What is it then?'
 'It could just as well be a cousin who got a job in Amsterdam and is now coming to stay with him until he finds a room?'
 'That's possible, but it doesn't seem likely to me.'
 'To always say that someone is homosexual right away,' she said. 'It doesn't seem like a homosexual to me at all. He is a man on his own, just like Frans Veen. You don't say that he is a homosexual, do you?'

This reader gave *De buurman* the highest score of 7 for both literary and general quality: ‘Use of language and difficult to put down despite being about little or nothing’ (ID_4074). It received two 6s from the following reader: ‘On the basis of the style, theme, and Voskuil’s other titles’ (ID_11019). There were also readers who judged it more sparingly. ‘The concept is literary, the style less so’ (ID_5976), wrote a participant who gave it a 5 for literary quality and a 6 for general quality. The last substantiation I want to quote is from a person who gave a 2 for literary quality and a 4 for general quality: ‘It is very much an I-story. Voskuil constantly repeats himself. The book is a homely continuation of *Het Bureau*. His style gets boring in the long run’ (ID_2300). It is very likely that the large amount of seemingly plainly styled dialogue is responsible for the outlying position of the novel in Figure 5.5.

Next, I compared the vocabulary of the group of novels awarded the highest literary quality with that of the group with the lowest scores, using the keyword function in Laurence Anthony’s AntConc as I did in earlier experiments. I excluded the middle group. In the novels with the highest scores, the following personal pronouns were used relatively more often than in those with the lowest scores: *hij* (he), *hem* (him), *haar* (her), *zij* (she/they, so either singular or plural), *me* (me), and *ge* (you). This last word *ge* is characteristic for the four Flemish novels in the selected books. The Literary novels with the lowest ratings featured significantly more occurrences of *we* (we) and *jullie* (you, plural). The highest-rated novels therefore seem to place a little more emphasis on the singular and the individual and the lowest-rated novels more on the plural and the collective. This pattern thus differs from that observed in the translated literary novels: there was more *ik* (I), *je* (you, singular), and *mij* (me) in the high-rated novels and more *hij* (he), *hem* (him), and *hen* (them) in the low-rated ones.

As for the translated novels, there is a difference in verb tenses: in the originally Dutch novels that were considered highly literary by the readers, the present tense seems to be more dominant, with relatively more frequent *zegt* (says), *vraagt* (asks), *kijkt* (looks), and *heeft* (has). This as opposed to more *was* (was), *waren* (were), *vroeg* (asked), *zei* (said), and *keek* (looked) in the least literary books. In Chapter 6 of this book, I will elaborate on the differences I have found per category for the literary novels written by women and men, translated and not translated.

A few other intriguing contrasts came to the fore. The novels that received the highest scores for literary quality contained the following adverbs and conjunctions relatively more often: *nog* (yet), *al* (already), *ineens* (suddenly), *telkens* (every time), *alsof* (as if), *zelfs* (even), *eens* (once), *zoals* (like), *als* (as if), *zonder* (without), and *zodra* (as soon as). A possible explanation for the conjunctions might be a more frequent use of compound sentences. Furthermore, the pronoun *die* (who/that) was used significantly more often – either as a relative pronoun (the man *who*...) or as a demonstrative pronoun (*that* man). The same goes for the indefinite article

een (a) in the lowest rated novels – in the translations it was the definite article *de* (the). It is unclear to which features of the novels this may relate. And a final riddle to mention: the Dutch-language novels by male authors with the lowest scores for literary quality more often contain *broeder* (brother), *moeder* (mother), *baby* (baby), and *mama* (mama) than the novels with the highest scores. I will return to this issue in Chapter 6 and Chapter 7.

The remainder of this chapter will first analyse how two novels by Kluun and Koch, which were published almost simultaneously in 2011, compare in terms of the opinions of the participants in The National Reader Survey and in terms of their style. Next, I will go into two novels by Arnon Grunberg, to try to explain why these novels ended up between Romance novels in the first experiment I presented in Chapter 2.

A Literary Duel: Kluun versus Koch

In the last week of January 2011 two different publishing houses presented two new novels by two well-known Dutch authors: Herman Koch, *Zomerhuis met zwembad* (*Summerhouse with Swimming Pool*) and Kluun, alias of Raymond van de Klundert, *Haantjes* (**Alpha males*). Both novels ended up in our list of 401 novels most sold from bookstores and most borrowed from public libraries in the years 2010–2012. Instigated by Kluun's publisher, the media paid a lot of attention to this almost simultaneous publication – 'It's called the literary duel K2', a literary reviewer wrote (Fortuin). In this section I will show how participants in The National Reader Survey rated the two novels and how they motivated their scores, and what a computational analysis of the texts themselves may uncover (cf. Van Dalen-Oskam 2016). First, I will introduce the two authors in more detail.

Both Kluun and Koch earlier wrote novels that in the Dutch language-area brought them a comparably large success in sales, and were widely translated. They did not bring them an equal amount of literary acclaim, however. Kluun entered the bestseller list in 2003 with his debut novel *Komt een vrouw bij de dokter* and readers (and non-readers) couldn't stop talking about it. Reviewers found the book vulgar and lacking any literary quality, but the reading public certainly did not shy away from the novel. Carmen, the wife of main character Stijn, is ill and ultimately dies from cancer. Stijn supports her as much as he can, but also increasingly seeks out sex with other women, which many readers found shocking. At the time of writing this chapter (March 2022) according to the translation database of the Letterenfonds (Dutch Foundation for Literature) the novel was translated into 25 languages. In 2007 it was translated by Shaun Whiteside and published in English as *Love Life* by Ray Kluun.

Herman Koch's big breakthrough came in 2009, with his best-selling novel *Het diner*, which was also in our corpus of 401 novels. By March 2022, *Het diner* was translated into 38 languages. Sam Garrett was responsible for *The Dinner*, the English translation that was first published in 2012. This book, which has also been turned into several movies with the same title, is about two couples who try to come to terms with the knowledge that their teenage sons have committed a serious crime.

While on an international level Koch is mainly known as the author of literary fiction, the Dutch audience first got to know him as an actor and comedian in *Jiskefet*, a humorous, absurdist-satirical series (1990–2005) of Dutch public broadcaster VPRO television that in some ways may be compared with the humour of Monty Python's Flying Circus. This is something to keep in mind when reading motivations of ratings of Koch's novels.

Other works of both Koch and Kluun have also been translated, but Koch's work is spread more widely than that of Kluun. A total of eight of Koch's works have appeared in Italian, six titles are available in German, Finnish, Hebrew, and Russian, five in French, and four in Bulgarian, Danish, English, and Spanish. Koch's next novel after *Het diner* was *Zomerhuis met zwembad* (*Summerhouse with Swimming Pool*), as of March 2022 translated into 25 languages, clearly riding the wave of the success of *Het diner*. The global distribution of Kluun's work is topped by four titles translated into Bulgarian, three into Czech, two into Complex Chinese (for Taiwan) and one into Chinese (published in Wuhan, in mainland China). Kluun's *Haantjes* (**Alpha males*) has until now only been translated into one language; it was published in Complex Chinese in 2012.

Haantjes (**Alpha males*) is all about Stijn van Diepen's and Frenk Versteeg's creative and strategic marketing agency Merk in Uitvoering – the name plays with the Dutch phrase *werk in uitvoering* (work in progress), so it signals something like 'brand in progress'. Their one-year-old company is not making much progress in terms of assignments and income. They present themselves as new and hot and very much in demand, but according to their secretary Maud they mainly produce reports. Then, Frenk's gay neighbour, the international model Charles, has a great idea. Soon the Gay Games will be held in Amsterdam. That will attract visitors from all over the world. And gays are big spenders. What if we replaced the colour red with pink in all the flags? And so the project Gay Flags is born. Stijn can already see the money coming in and also invests in the project himself. He informs his wife Carmen, nominated for the title Business Woman of the Year, only after he has put his money on the table. Stijn would have done better to involve her in the project much earlier...

On Sunday 23 January 2011 the Dutch left-leaning quality newspaper *de Volkskrant* published an unexpectedly favourable review of *Haantjes* by literary reviewer Arjan Peters. Peters explicitly stated that he was negatively prejudiced when he

started reading, mainly because of the vulgarity of Kluun's earlier novels, but that all his objections disappeared while reading. Another influential literary reviewer, Arjen Fortuin, wrote extensively about both authors and both novels on 28 January 2011 in the quality newspaper *NRC-next*, comparing the novels with each other in both print runs and literary quality. Fortuin was also positive about *Haantjes*. He agreed with Peters that Kluun's novel was moving away from his earlier 'kitsch' and suggested that Koch took the opposite route; from being an author with a specific literary audience in mind for high-quality literary novels, he became a writer for a much wider audience when he hit the jackpot with *Het diner* (*The Dinner*). Fortuin emphasized that although both Kluun and Koch have very different backgrounds, the print runs of their new novels were more or less in line with eighty thousand copies for Kluun and one hundred thousand for Koch. According to Fortuin, another similarity is that both novels in fact are 'men's books about being a man'.

The main character in Herman Koch's *Zomerhuis met zwembad* (*Summerhouse with Swimming Pool*) is general practitioner Marc Schlosser, married to Caroline and father of two beautiful daughters aged eleven and thirteen. Marc works in Amsterdam and he does things differently from other GPs. He spends twenty minutes per consultation – double of what other GPs do. Personal attention is his trademark, and it works. His patients are just as exclusive and include many celebrities. When actor Ralph Meier drops by without an appointment, this leads to further personal contact. Ralph invites Marc and his family to a barbecue, where they get to know Ralph's wife Judith and their two sons. This leads to an invitation to visit the summer house Ralph and his wife have rented on the French coast during the next holiday. Feelings of discomfort that have arisen in the meantime come to fruition there, with irreversible consequences for Marc and his family. Literary critic Arjen Fortuin, in his piece for *NRC-next*, described the novel as a smooth read, funny, and not superficial, but all in all more of the same when compared to *Het diner*.

Fortuin was somewhat disappointed with Koch's *Zomerhuis met zwembad*. And although he was positive about *Haantjes*, he thought the literary battle was unequal: Fortuin rated Kluun as 'playing a couple of divisions lower' and marked *Haantjes* as very lightweight. To him, Koch seemed to have the best chance of winning both the literary duel and the overall number of sales.

When we rank our corpus of 401 books according to number of sales and borrowings, the two novels ended up not very far apart: Koch's novel at rank twelve and Kluun's at seventeen. The complete list can be found on the companion website to this book (see Appendix 3) or using the litRiddle R package, sorting the books by their ID number.

In The National Reader Survey, Koch's novel received considerably more opinions than Kluun's book (714 versus 208). Figures 5.6 and 5.7 show that Koch mainly

received scores of 5 or 6 for literary quality and that the opinions about Kluun’s book are much more evenly spread across scores 1 to 5. In the readers’ eyes, Koch clearly won the literary duel and Kluun definitely was the loser. They were less positively surprised by the new Kluun than the two literary reviewers mentioned earlier.

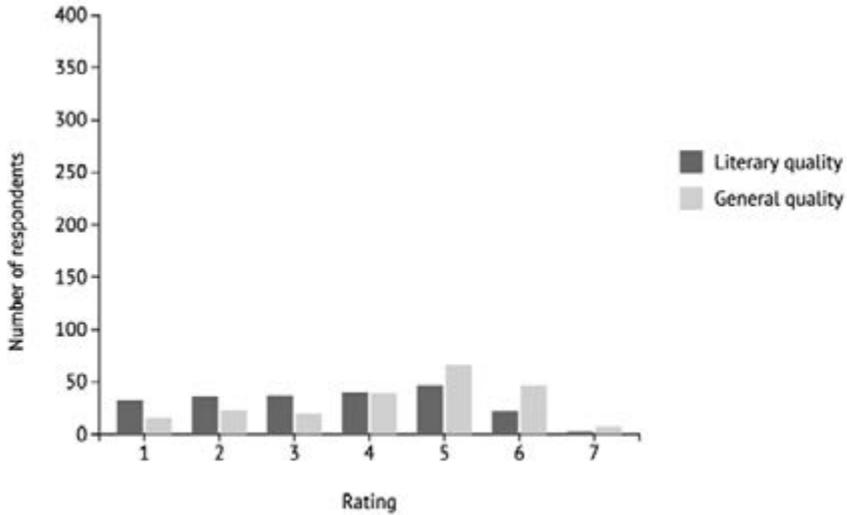


Figure 5.6: Ratings of Kluun, *Haantjes* (*Alpha Males)

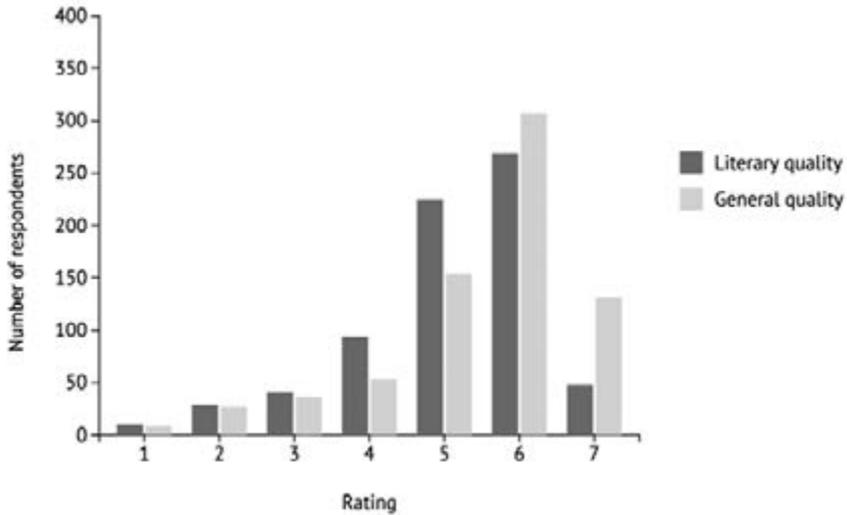


Figure 5.7: Ratings of Herman Koch, *Zomerhuis met zwembad* (Summerhouse with Swimming Pool)

To gain some insight into this literary duel, I present a representative selection from the justifications readers gave for their scores. The ratings are indicated before the quote, from score (1), definitely not literary, to (7), highly literary.

(1) **Kluun:** Poorly written, within three pages it is clear what is going to happen, very flat/vulgar characters (ID_4638).

(1) **Kluun:** The story is based on the book *Kaas* [*Cheese*, by Willem Elsschot], so not very creative. Predictable storyline, not much depth (ID_1497).

(1) **Kluun:** Spun-out anecdote not strong stylistically (ID_5961).

(1) **Koch:** In my opinion, *Zomerhuis met Zwembad* [*Summerhouse with Swimming Pool*] is a typical example of a book written according to a certain concept. A book that was written purely to achieve a high position in the bestseller list. For me, literature is a text that contains an original voice, the drive to tell a story, originality in ideas, style, and use of language (ID_10693).

(2) **Kluun:** *Haantjes* by Kluun has a bad story and is overrated by the public (ID_1575).

(2) **Koch:** Plot is too simple and over the top. Popular (ID_12306).

(3) **Kluun:** I liked *Komt een vrouw bij de dokter* [*Love Life*]. *Haantjes* [**Alpha Males*] bored me (ID_4958).

(3) **Koch:** I don't like the style, the content is too predictable, and the characters lack depth (ID_6032).

(4) **Kluun:** I like the way the book is written but I think the content is not very high brow (ID_3485).

(4) **Koch:** Sometimes somewhat cheap effects and insufficient literary distance (ID_658).

(5) **Kluun:** Badly written, probably meant to be literature, but does not meet my (subjective) standard regarding the quality of a novel (ID_11529).

(5) **Koch:** Thinking required (ID_1824).

(5) **Koch:** I don't find that easy to answer. I like to read a good story. But the intensity of it, the impression the book leaves on me, is certainly important. A feeling of being part of the book. Well, difficult question (ID_10245).

(6) **Koch:** Good story, developed with deeper thoughts. Interesting people who have to make choices in difficult situations (ID_7500).

(6) **Koch:** Difficult, but in my opinion it is a stayer (ID_2745).

(7) **Koch:** The possibility to identify and the recognizability of the character are developed to such an extent that the humour becomes almost visual (ID_9251).

(7) **Koch:** Linguistically good. Tension is built up well, very humorously written. Good story line. Continuously captivating and interesting (ID_7844).

The motivations show that Kluun's novel was considered too anecdotal and popular in style, whereas Koch's was perceived as having a literary style and deeper layers.

Several motivations are strongly reminiscent of those found in the previously described reactions in the media. It is certainly not unthinkable that literary reviewers were actually responsible for some of the motivations above. We do know that some participants of The National Reader Survey actually were literary reviewers, because some of them explicitly mentioned this in their motivation of a literary rating.

To get a clearer view on the differences between the two novels themselves, I compared them with four other books, two with high scores and two with relatively low scores for literary quality: the literary top scorers *Alsof het voorbij is* (*The Sense of an Ending*) by Julian Barnes and *Godenslaap* (*While the Gods Were Sleeping*) by Erwin Mortier, and the much less literary *Shopaholic & Baby* (*Shopaholic and Baby*) by Sophie Kinsella and *Nooit ziek geweest* (**Never Been Ill*) by Nico Dijkshoorn. Before you take a look at Figure 5.8, have a taste of the style of the novels and predict the outcome... The first quote is from Chapter 17 of Part I of Kluun's *Haantjes* (in my own translation). The second is from Chapter 11 of Koch's *Zomerhuis met zwembad* in Sam Garrett's translation (cf. Van Dalen-Oskam [2021] 223–224).

Maud came in with the boy. He introduced himself as Hylke Strijpstra and apologized for his father's last-minute absence ('a big order in Dubai'). Because of the bad weather he was a bit late.

Frenk said that he wanted to save the part of the presentation of Merk in *Uitvoering* until next time, when father Strijpstra would be present. He wanted to use this conversation to gather the information we needed to make the right decision. Mr Strijpstra's youngest son embarked on his story. The Dokkumer Vlaggen centrale [The Flag Factory in Dokkum] had existed since 1797. They were the largest supplier in Western Europe. Of all the flags that we saw in the Netherlands, both commercial and private, over three quarters came from the Dokkumer Vlaggen centrale. Even the large champion's flags hanging from the ridge of the Arena came from their company.

With a wave of his hand Frenk silenced Mr Strijpstra jr. Frenk believed that the Dokkumer Vlaggen centrale could make flags like no other. What else did they supply apart from flags?

'Daddy...' I felt a hand on my elbow and turned to look. There stood Julia with the languid boy who had shaken my hand earlier, but who I had already forgotten was called Alex. Standing slightly behind them were two other boys and two girls. 'Can we go out to get some ice cream?' she asked. 'It's really close.'

In terms of timing it was both a good and a bad moment. There was a chance that the slight sultry edge to our – superficially – innocent conversation about teenage bedrooms, baby seal posters, and horse books might be lost for good.

On the other hand, here I stood with my thirteen-year-old daughter, living proof that this charming man – me – was capable of siring a child. And not just any child, but a dreamy-eyed blonde who threw fifteen-year-old boys' hormones into overdrive the moment they saw her. I won't try to deny it: I take pleasure in being with my daughters in places where everyone can see us together.

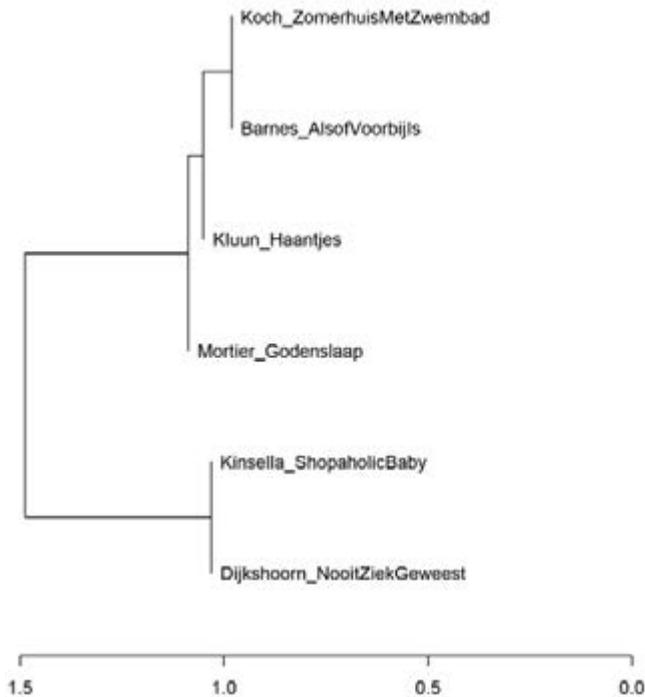


Figure 5.8: Kluun, *Haantjes* (**Alpha Males*), versus Koch, *Zomerhuis met zwembad* (*Summerhouse with Swimming Pool*), Cluster Analysis (1000 most frequent words).

Compared with Barnes, *Als of het voorbij is* (*The Sense of an Ending*), Kinsella, *Shopaholic en baby* (*Shopaholic and Baby*), Dijkshoorn, *Nooit ziek geweest* (**Never Been Ill*), and Mortier, *Godenslaap* (*While the Gods Were Sleeping*).

The cluster analysis of which the visualization is shown in Figure 5.8 was done with the Stylo package for R and uses as input the thousand most frequent words in all six novels. The further to the right books are connected by a vertical line, the more similar they are in word use and word frequencies. The result is that *Haantjes* and *Zomerhuis met zwembad* are much less distant from each other than readers' opinions would suggest. They are both close to the novels *Als of het voorbij is* by Julian Barnes and *Godenslaap* by Erwin Mortier, which are highly regarded for

their literary quality, and in this cluster of four clearly distinguish themselves from *Shopaholic & baby* by Sophie Kinsella and *Nooit ziek geweest* by Nico Dijkshoorn, which are considered to be far less literary. The measurements in this small selection of novels may support the idea that Kluun's novel was undervalued by readers. So the unexpectedly positive reviews that *Haantjes* received in prestigious newspapers such as *NRC-next* and *de Volkskrant* did not boost Kluun's low literary prestige.

Grunberg and Romance

Arnon Grunberg has been one of the most prolific and visible literary authors in the Netherlands since his debut novel *Blauwe maandagen* (*Blue Mondays*) appeared in 1994. He churns out novel after novel, writes newspaper columns, and is visible on television and at all kinds of literary events. Grunberg has received many literary awards for individual novels and in 2011 was awarded the prestigious Constantijn Huygens oeuvre prize, followed in 2022 by the equally prestigious P.C. Hooft oeuvre prize for prose fiction. In March 2022 the translation database of the Letterenfonds (Dutch Foundation for Literature) mentioned translations of Grunberg's work into thirty languages. Topping the list was German, with a total of eighteen translated titles, followed by French (11), Hungarian (11), English (10), Italian (9), Czech (8), Portuguese (7) and Turkish (6). Grunberg's internationally most well-known novel was *Tirza* (2006), translated into 21 languages. Sam Garrett was responsible for the English version, *Tirza*, published in 2013. The two novels in our corpus have not been translated into English yet but are available in four (*Huid en haar* – **Hair and Hide*) and six (*De man zonder ziekte* – **The Man Without Illness*) other languages than Dutch.

In Chapter 3 I showed how Grunberg's two novels end up between the Romance novels when I compared the most literary and least literary novels in our corpus using text analysis methods. In this section I will try to explain this. To give you a taster, I will start with a quote (cf. Van Dalen-Oskam [2021] 226).

'What are you waiting for?' asks Lea.

She is wearing a black wool coat with a fur collar, bought second-hand. She cannot afford such coats new.

Lea travels lightly. One backpack is all she needs for five days. A hairdryer gets most of the wrinkles out of your clothes.

On her knee is a hand. But a hand on a knee is not intimacy.

'What exactly are you a connoisseur of?' A professor had asked her during a standing reception that evening, seemingly casually touching her upper arm. She had found it unpleasant. The question and the touch.

An hour before, she had hung her dress in the bathroom on the rail to which the shower curtain was attached and treated it with the hair dryer. The wrinkles did not come out as well as she had hoped. Tomorrow morning she will go home, then she can have her dress dry-cleaned.

These are the first lines – in my tentative translation – of *Huid en haar* (*Hair and Hide) by Arnon Grunberg. Roland Oberstein is writing a book about the economic bubble, and has a pretty complicated life. He is divorced and his son lives with his ex-wife, he has a girlfriend in Amsterdam, and currently works in Fairfax, in the United States. At a conference in Germany on the Holocaust – a subject he says he also finds interesting from an economic perspective – he meets the American scholar Lea, who is writing a book about Rudolf Höss, commandant of Auschwitz. Lea's marriage has been without sexual intimacy for some time. For the first time since ages she is looking for a new sex partner and manages to seduce Roland. Back in the Netherlands, Roland allows himself to be seduced by a female student. Meanwhile, Lea's husband has fallen fatally in love with an illegal Guatemalan male migrant, whom he forces into sex under the promise of a green card for him, his wife, and his child. The entanglements between all these characters become more and more complicated and ultimately end badly for some of them.

A participant wrote about *Huid en haar*, giving the novel the highest score, 7, for both general quality and literary quality: 'Grunberg knows how to draw you in and make you feel disgusted (about humanity?) in a way that is extremely clever. It does not make you happy' (ID_6868). The next reader gave the same rating and motivated it in these words: 'A lot of depth, also in terms of characters. Fascinating, very creative writing style. A game with language/structure/characters' (ID_4039). Another participant also emphasizes the importance of style, scoring 5 for general quality and 6 for literary quality: 'Because of the writing style. Grunberg does not necessarily accommodate the reader, he does not write a simple entertaining story' (ID_3363). There are very few exceptions to the praise. 'I think it's lazy writing' (ID_9814), wrote a reader who gave the novel a 4 for general quality and a 5 for literary quality.

Grunberg's other novel in our corpus, *De man zonder ziekte* (*The Man Without Illness), is about young Samarendra – usually 'Sam' – Ambani, who was born and raised in Switzerland, has a Swiss mother and a father who comes from India. He has a girlfriend, Nina, and works as an architect. He took part in a competition to design an opera house in Baghdad and was among the three finalists. At the invitation of Hamid Shakir Mahmoud, the man who organized the competition, Sam travels to Baghdad. Mahmoud personally guarantees Sam's safety and Sam is looking forward to the meeting – which never happens. In Iraq, Sam has some gruesome experiences and his life completely changes, but his calling to design

buildings does not disappear. And when he has to go to Dubai for a new project, he even looks forward to it. That visit ends even worse, however. This novel begins as follows, again in my translation (cf. Van Dalen-Oskam [2021] 227):

For his trip, Samarendra Ambani and his girlfriend bought a new suitcase. A grey suitcase with wheels. Because Samarendra's girlfriend was afraid that he would not recognize his suitcase on the luggage belt, she tied a light green hair ribbon around the handle.

He would have preferred not to have that hair ribbon. Samarendra, who is called Sam by most people, likes to come across as a professional traveller, someone who has been almost everywhere and therefore feels at home almost everywhere. Such a hair ribbon was more suited to a middle-aged tourist with homesickness and barely suppressed fears. But he did not want to disappoint his girlfriend. She had said: 'Then you will always think of me when you see your suitcase.'

From the substantiations of ratings of *De man zonder ziekte* I quote two that mention the writing style in different ways. This reader gave the novel a 7 for both general and literary quality: 'In this book I found his writing style very compact, he tells the story without unnecessary digressions, it was an easy read' (ID_12100). The following somewhat longer motivation also refers to Grunberg's use of language: 'The structure, plot, and outcome of the story, and the suspense, the use of language, the contrasts between the main character, his personality, and the situation he finds himself in are described in such a way that you constantly want him to reconsider things. It keeps you active, you want to write along, to warn. I think it is literary if you can use language to bring that about in your reader' (ID_468). This reader rated the general quality with a top score of 7 and the literary quality with 6. Also for this novel very few readers were not enthusiastic at all. The following participant gave the book a 1 for general quality and a 3 for literary quality: 'The story is utter rubbish. A seriously overrated novel' (ID_5671).

Readers found the style of Arnon Grunberg's two novels easy or enjoyable to read, found the use of language creative and compact, and appreciated this positively – these were the arguments for a high score for literary quality. None of the Grunberg readers who took part in The National Reader Survey and shared their opinions about one of his novels referred to Romance. The software did, as I already pointed out. Why? For a taste of this genre I again selected *Shopaholic en baby* (*Shopaholic and Baby*) by Sophie Kinsella that I used in a previous experiment in this chapter. The Dutch translation was produced by Mariëtte van Gelder. I quote the first paragraphs from the first chapter of the English original to give an impression of the writing style. Rebecca Brandon is pregnant and has absolutely no idea what to expect:

OK. Don't panic. Everything's going to be fine. Of course it is.

Of *course* it is.

'If you could lift up your top, Mrs Brandon?' The sonographer has a pleasant, professional air as she looks down at me. 'I need to apply some jelly to your abdomen before we start the scan.'

'Absolutely!' I say without moving a muscle. 'The thing is, I'm just a teeny bit... nervous.'

I'm lying on a bed at the Chelsea and Westminster hospital, tensed up with anticipation. Any minute now, Luke and I will see our baby on the screen for the first time since it was just a teeny blob. I still can't quite believe it. In fact, I still haven't quite got over the fact that I'm pregnant. In nineteen weeks' time, I, Becky Brandon, née Bloomwood... am going to be a mother. A *mother*!

Luke's my husband, by the way. We've been married for just over a year and this is a one hundred per cent genuine, honeymoon baby! We travelled loads on our honeymoon, but I've pretty much worked out that the baby was conceived when we were staying in this gorgeous resort in Sri Lanka called Unawatuna, all orchids and bamboo trees and beautiful views.

Becky works as a personal shopper at a new prestigious department store, The Look, which is on the brink of collapse due to a lack of business/patrons, so she has all the time in the world to shop privately, looking for the perfect baby clothes, bedding, cots, pushchairs, and whatnot. Everything has to be of the highest fashion standard – and that includes the person who will guide her through the pregnancy and deliver the baby. As soon as Becky hears of Venetia Carter, 'the must-have celebrity obstetrician', she knows what she wants. But not everything in Becky's pregnant life goes according to plan. Yet with her creativity and personality, she manages to solve problems in an impressive way.

Here too, I quote a couple of substantiations written by readers who rated the book. 'It is no literary masterpiece, easily written, but with a lot of humour' (ID_13068) wrote a reader who gave the novel a 4 for both general and literary quality. Another one gave the book a 6 for general quality and a 2 for literary quality and goes a little deeper into the use of language than the previous participant: 'It is written as an easy read, so easy words, sentences not too long. That's not a bad thing at all, but not literary' (ID_8052). A final quote, from a reader who gives the novel a 3 for overall quality and a 1 for literary quality: 'Too many colloquialisms, implausible events, too many "easy" storylines that are predictable, chewed-out explanations.' The extensive motivation continues: 'This book is one of a series and all the books are structured the same, but with a different execution. No multiple layers in the book. What you read happens, no more and no less. So, not very original and just plain reading' (ID_12557).

The style of *Shopaholic en baby* is repeatedly described as easy: easy words, easy to read, shortish sentences, simple language, no distinctive writing style, and in the last motivation I quoted, even more concrete, too much spoken language to be considered literary. The story is easy and pleasurable to follow, but does not linger and does not inspire deeper thought. Therefore most readers considered the novel not very literary. Only one of the respondents who shared a motivation disagrees – an exception at the other end of the spectrum than in the motivations for the reviews of Arnon Grunberg's novels. She gave *Shopaholic en baby* the second-highest score, a 6, for both general and literary quality, and substantiated this opinion as follows (I present a literal translation): 'It is a book that many ordinary women grapple with, although it is dramatized' (ID_2558).

In 2018, Yra van Dijk published her monograph *Afgrond zonder vangnet. Liefde en geweld in het werk van Arnon Grunberg* (**Abyss Without a Safety Net: Love and Violence in the Work of Arnon Grunberg*). Van Dijk presents a penetrating analysis of Grunberg's work, also paying attention to his personal history as a child of Jewish parents who were both deeply scarred by the Second World War. This background proved to be of great importance when analysing the nature and content of everything Grunberg wrote until 2018. His novels are not particularly mild, Van Dijk writes. 'Grunberg's protagonists are invariably unlucky people with nothing but good intentions, who at the end of the story have destroyed everything they loved with their own hands' (13). Characteristic themes of his work are care, love, illness and violence, satire, parody, and social engagement.

At various points in her book, Van Dijk refers to Grunberg's 'transparent style'. She links this to his productivity as a writer. According to her, Grunberg obsessively wants to be read by as many readers as possible (15). His works are so transparent in style because of the vital importance of communication with the reader (14). Van Dijk does not explicitly describe what exactly makes Grunberg's writing style transparent. However, she does present several observations that help put the style more in perspective. For *Huid en haar* (**Hair and Hide*) she points out that despite the transparent style, the novel is quite complex. It is inherently unstructured. The perspective is constantly changing; the story continuously shifting in focalizer. Van Dijk suggests that the novel's structure thus illustrates the confused reality of protagonist Roland Oberstein's life (204). Furthermore, the book is full of trivial details and banal conversations, which, according to Van Dijk, is exactly what we would *not* expect in a literary novel. She suggests that the goal of this abundance of details might be 'a reality effect'. Precisely by presenting ever more details, the text can claim transparency (208). 'From nappies to lice: the banal which is usually found in the margins of literary narratives has been pushed to the fore here' (211).

In *Huid en haar*, Van Dijk identifies a tension between scientific and social work on the one hand, and the human duty of care on the other – Roland Oberstein finds

it particularly difficult to comply with his role of caregiver. Van Dijk suggests the double bind in Grunberg's stories is that 'the ego of his protagonists is damaged by caring for the other, but not caring is also impossible for them' (212). This ethical dilemma is made visible exactly by the extreme realism of the text, Van Dijk argues (213). The comprehensibility and transparency that are Grunberg's trademarks, writes Van Dijk, is a false transparency 'because it is a reality that cannot signify anything. It is precisely in this *non-transparent realism* where the ethics of this text lie. At this level, it does make things difficult for the reader' (215). In my opinion, Van Dijk argues that the 'transparent style' – although easy to read – precisely emphasizes the unusually heavy underlying ethical problems in the novel. The simple style, I deduce, can be seen as a means by which the content is shaped all the more concisely. The reader experiences the content as harder and more confronting because of the simple writing style.

But also in terms of content, Grunberg makes use of story elements that are unusual for literary novels. Van Dijk takes as example the dramatic ending of *Huid en haar*; the fact that the student who seduced Oberstein seriously injures herself when she falls from her horse, 'is a nineteenth-century conclusion to a twenty-first-century soap opera' (210). Only occasionally, a participant in The National Reader Survey used the word 'soap opera' or 'soap' – but not when writing about a book by Grunberg. Sophie Kinsella, however, did not escape this fate. One reader wrote to substantiate the very lowest score for literary quality of Kinsella's *Shopaholic en baby*: 'This is a flat, soap opera-like story without any depth or layers' (ID_6850).

And so it becomes clearer why Grunberg's novels appeared among the Romance novels in the measurements I did: the use of language, possibly the prominence of colloquialisms, and perhaps an emphasis on words to do with care and relationships – all obvious parallels with Romance novels, with their thematic emphasis on everyday things and social processes. In terms of content, however, the differences are clear: whereas books labelled Romance invariably have a happy ending, the same cannot be said of Grunberg's novels.

In an experiment similar to the one in the previous section I compared Arnon Grunberg's two novels with two novels by Sophie Kinsella, which received very low scores. I added two other novels from the extremes of the overall literary rankings: the lowly-rated *Geknipt voor jou!* (**Perfect for You!*) by Chantal van Gastel and *Nooit ziek geweest* (**Never Been Ill*) by Nico Dijkshoorn, and the highly literary *Alsof het voorbij is* (*The Sense of an Ending*) by Julian Barnes and *Godenslaap* (*While the Gods Were Sleeping*) by Erwin Mortier. Based on the scores for literary quality, we would expect Grunberg's novels to be more like the books by Barnes and Mortier, but Figure 5.9 proves this to be a wrong assumption. Grunberg's two novels are slightly more like the four least literary novels in the selection. Once again, as a reminder: this only tells us something about the selected novels and how they compare to each

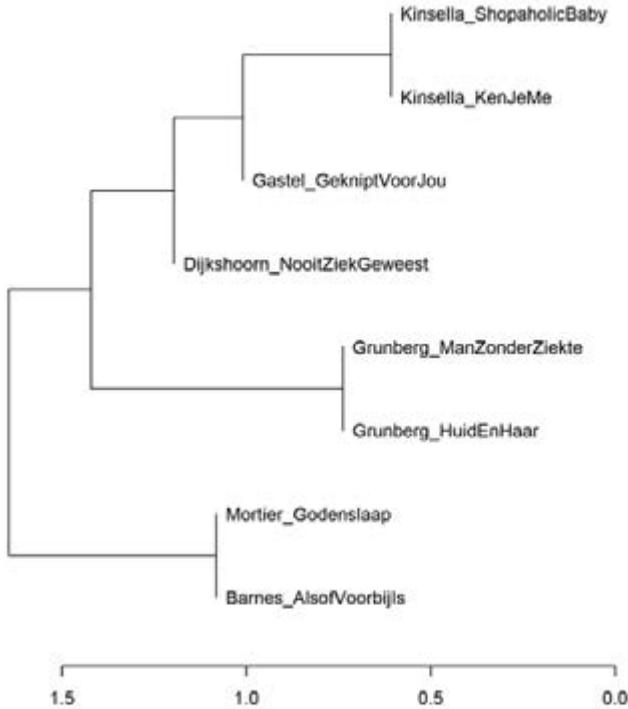


Figure 5.9: Grunberg, *Huid en haar* (*Hair and Hide) and *De man zonder ziekte* (*The Man without Illness), versus Kinsella, *Shopaholic en baby* (*Shopaholic and Baby*) and *Ken je me nog?* (*Remember me?*), Cluster Analysis (1000 most frequent words).

Compared with Barnes, *Als of het voorbij is* (*The Sense of an Ending*), Van Gastel, *Geknipt voor jou!* (*Perfect for You), Dijkshoorn, *Nooit ziek geweest* (*Never Been Ill), and Mortier, *Godenslaap* (*While the Gods Were Sleeping*).

other. As soon as more or other texts are used, the outcome changes. Nevertheless, this approach is useful to get an overall impression of the selected texts. It does not say anything about the place of a novel compared to all other published novels. These additional experiments highlight the exceptional style – for literary novels – of Grunberg’s works. This would not have become apparent had the Romance genre been excluded from the research.

Conclusion

I started this chapter with an overview of the place in the overall ranking of Literary novels written by male authors. The number 1, *Als of het voorbij is* (*The Sense of an Ending*) by Julian Barnes, was discussed in a bit more detail to get a sense of readers’ opinions and their motivation. Next, I analysed the forty translations on the list

with a principal components analysis (PCA). The PCA yielded some outliers, two of which may be explained by being translations of much older novels than the rest. On the whole, however, the results of the PCA did not present any clear differences between the novels with the highest ratings for literary quality and those with the lowest scores or with the group in between.

A keyword analysis uncovered that the novels with the highest literary quality had a slightly higher emphasis on a first-person narrator and/or contained more dialogue, and seemed to be set in the present tense more often. The least valued books were suggested to have a third-person narrative more often and be set in the past tense. Furthermore, for the most highly rated group, the use of a number of conjunctions stood out. These included the archaic *doch* (yet), which was mainly reserved for David Mitchell's, *De niet verhoorde gebeden van Jacob de Zoet* (*The Thousand Autumns of Jacob de Zoet*), and could be recognized as a conscious stylistic choice of the translators. For the fifty Dutch originals, the same comparison showed that the novels with the highest scores for literary quality emphasized the singular and the individual whereas those with the lowest scores focused more on the plural and togetherness. Translations and originals shared a comparable difference in verb tenses: in novels perceived as highly literary by the readers the present tense seemed to be more dominant, while the past tense was more visible in the group with the lowest scores.

Several of the fifty Dutch Literary novels were discussed in more detail. I started with the novels ranked 3 and 4, both by Flemish authors: Erwin Mortier's *Gestameld liedboek* (*Stammered Songbook*) and Tom Lanoye's *Sprakeloos* (*Speechless*). After presenting some readers' opinions about these two books I added some motivations about the novel that received the lowest score for literary quality, Mart Smeets' *De afrekening* (**The Reckoning*). I hypothesized that Smeets' reputation as a sports journalist, writing novels about popular sports such as cycling, biased readers' opinions.

A comparable observation came from a discussion of two novels that were published at the same time and as such were much commented on in Dutch media: Herman Koch's *Zomerhuis met zwembad* (*Summerhouse with Swimming Pool*) and Kluun's *Haantjes* (**Alpha Males*). In an experiment comparing both novels with two highly rated and two lowly valued novels, both ended up closer to the ones with high scores for literary quality than those with low scores. This suggested that Kluun's novel may have been undervalued by readers. Kluun's low literary prestige based on his earlier works did not get a boost by the positive reviews that *Haantjes* received in prestigious newspapers.

Finally, I addressed the observation from Chapter 3 that two highly valued novels by Arnon Grunberg seemed to be more adjacent to the least valued Romance novels than to other Literary novels with high scores for literary quality. I discussed

the novels in some more detail, referring to Yra van Dijk's 2018 monograph about Grunberg's work. Van Dijk observed that Grunberg's style is 'transparent', but did not substantiate this with textual examples. My discussion of reader opinions and of text analysis experiments, however, substantiated her observation and made it much more concrete. It became clear how a similar style in Grunberg's novels and in Romance novels like *Shopaholic en baby* can inform readers making totally different judgements about the literary quality of those novels.

This once again emphasized that there must be other things at play than relatively easily measurable linguistic characteristics. An overview of the results of more high-level measurements is one of the main topics of the next chapter of this monograph. It will also address the gender issue in more detail and present a final experiment that brings Dutch Literary novels written by female and male authors together again.

References

All hyperlinks were last accessed in February 2023.

- Anthony, Laurence. *AntConc*. Version 3.5.8., Tokyo: Waseda University, 2019, <https://www.laurenceanthony.net/software.html>.
- Eder, Maciej, et al. 'Stylometry with R: A Package for Computational Text Analysis.' *The R Journal*, vol. 8, no. 1, 2016, pp. 107–121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.
- Fortuin, Arjen. 'Mogen wij nog haantjes zijn? Tweestrijd van de moderne man is het gemeenschappelijke thema van Kluun en Koch' ['Are We Still Allowed to Be Alpha Males? The Inner Turmoil of the Modern Man is the Communal Topic of Kluun and Koch', not available in English]. *nrc-next*, 28 January 2011, pp. 4–5, <https://www.nrc.nl/nieuws/2011/01/28/mogen-wij-nog-haantjes-zijn-11991928-a1367099>.
- 'Letterenfonds Vertalingendatabase' ['Dutch Foundation for Literature Translation Database']. *Dutch Foundation for Literature*, <https://www.letterenfonds.nl/nl/vertalingendatabase>.
- Peters, Arjan. 'De nieuwe Kluun mag er wezen: Hoogstaand amusement in Haantjes van Kluun. Of: hoe Stijn van Diepen goud geld dacht te verdienen aan de Gay Games in Amsterdam.' ['The Latest Kluun Deserves Praise: Elevated Entertainment in *Alpha Males* by Kluun. Or: How Stijn van Diepen Schemed to Get Filthy Rich by Commercially Exploiting the Amsterdam Gay Games', not available in English]. *De Volkskrant*, 23 January 2011, <https://www.volkskrant.nl/cultuur-media/de-nieuwe-kluun-mag-er-wezen~bof5a268/>.
- Van Dalen-Oskam, Karina. 'A Literary Rat-Race.' *Digital Humanities Conference 2016*, Kraków, 11–16 July 2016. Conference Abstract. <https://www.dh2016.adho.org/abstracts/267>.

- Van Dalen-Oskam, Karina. *Het raadsel literatuur: Is literaire kwaliteit meetbaar?* [*The Riddle of Literature: Can We Measure Literary Quality?*] Amsterdam University Press, 2021, <https://karinavdo.github.io/RaadselLiteratuur/>.
- Van Dijk, Yra. *Afgrond zonder vangnet: Liefde en geweld in het werk van Arnon Grunberg* [*Abyss Without a Safety Net: Love and Violence in the Work of Arnon Grunberg*, not available in English]. Nijgh and Van Ditmar, 2018.
- Verstraeten, Pieter. 'Na de race: Sport en literatuur rond 1900' ['After the Race: Sport and Literature Around 1900']. *Uit de marge: Kanttekeningen bij de cultuurhistorische canon* [*From the Margin: Side Notes to the Socio-Historical Canon*, not available in English], edited by Lizet Duyvendak and Jan Oosterholt. Literatoren, 2018, pp. 207–2013.

6. Style, Gender, and Genre

Combining Multiple Perspectives

Abstract: This chapter sums up which textual features beyond words were analysed and how they may play a role in readers' evaluation of literary quality. It discusses sentence structures, parts of speech (PoS), bigrams, and cliché expressions. It describes how topic modelling was used to analyse thematic differences and how predictive modelling was applied to explore which textual and contextual features most influenced the ratings for literary quality. Mean sentence length, in number of words per sentence, proved to be a very good indicator. Next, Dutch Literary novels by women and men are compared in several computational ways and the influence of the author's gender on readers' perceptions is described in more detail. Finally, the feedback loop of genre is discussed.

Keywords: Literary quality, Topic Modelling, Predictive Modelling, Sentence length, Computational Literary Studies, Distant Reading

Beyond Words

In the previous chapters, I combined the analysis of readers' ratings and motivations with different measurements of the most frequent words in the rated novels themselves. The results of these bag-of-words kind of measurements led to several intriguing observations, but all in all did not yield enough information to get a clear answer to the question how novels perceived as highly literary differed from those with less attributed value. I will start this chapter with an overview of the results of linguistic analyses that went beyond word frequencies. Here, Andreas van Cranenburgh played a key role. His PhD thesis *Rich Statistical Parsing and Literary Language* and other publications he co-authored addressed a diverse set of high-level linguistic analyses of the corpus of 401 novels.¹ I will then continue with a closer

¹ Because the page numbers in van Cranenburgh's thesis as it was defended (<http://hdl.handle.net/11245/1.543163>) generally do not agree with the page numbers of the slightly longer revised version

analysis of the gender bias we found. First, to finalize my own experiments, I will bring together the main results from the previous chapters in a comparison of the Dutch-language Literary novels by female and male writers. Next, I will present an overview of Corina Koolen's work in her PhD thesis *Reading Beyond the Female: The Relationship Between Perception of Author Gender and Literary Quality*. In my review of the two PhD-theses and other publications resulting from the Riddle project, I will focus on the main outcomes. Readers are referred to these publications for more detailed information about the underlying methods, statistics, and a review of relevant other publications that may help to better contextualize their scholarly work.² The chapter ends with a brief reflection on the feedback loop of genre.

Syntactical Analysis

Right after the start of the project The Riddle of Literary Quality in 2012, team members Kim Jautze, Corina Koolen, Andreas van Cranenburgh, and Hayco de Jong developed a first experiment focusing on syntactical analysis. Because we were in the process of compiling the Riddle corpus, they gathered a small pilot corpus of 32 novels, written in Dutch, from two categories: literary novels (general fiction) and what were then still marketed as so-called 'chick lit' novels – humorous-romantic books about young women in an urban environment pursuing a career and a relationship. (Some of these 32 novels later ended up in our corpus of 401 novels, where we labelled them as belonging to the Romance category.) Jautze et al. expected that many readers would have concrete ideas about differences between the two groups, for example, that chick lit novels are easier to read than literary novels. Their aim was to describe these differences at sentence level. They wanted to ascertain whether the proportion of certain sentence types differed in the selected novels, and whether they could confirm their intuition that the readability of the chick lit novels was related to a higher proportion of relatively easy sentence structures as compared to the literary novels. Easiest to read are simple sentences of one main clause, and then, in increasing difficulty, compound sentences with coordination of clauses, complex sentences with subordinate clauses, and finally a combination of complex and compound where, in compound sentences, subordinate clauses appear in different forms of complexity (Jautze et al. 72–74).

the author made available in his Github (http://andreasvc.github.io/phdthesis_v1.1.pdf) I refer to the chapter and chapter section in which the topic is addressed.

² Both PhD-theses and most of the publications are available online in open access and can be found through https://literaryquality.huygens.knaw.nl/?page_id=588. They are also directly available from the University of Amsterdam Repository, as cited in this book.

Compared with the sixteen chick lit novels, the sixteen literary novels appeared to contain significantly longer sentences, which means a higher probability of more complex sentences. They also had a significantly longer average word length (in number of characters), which may indicate the occurrence of slightly more difficult words – which are usually longer than easy words. The number of different words was also larger. A surprising observation was that it took the software considerably longer to parse the literary novels than the chick lit novels: the longer and more complex sentences of the literary novels for readers were also more difficult for the software than the sentences in the chick lit novels (75–76).

Jautze et al. also looked at smaller parts of sentences and the proportion of certain phrases and types of words. They found that the sixteen chick lit novels used more diminutives than the sixteen literary novels, and that the literary novels contained more prepositional phrases and more relative clauses. Both, they assumed, could indicate a larger share of description. All in all, the language use in the sixteen chick lit novels seemed to be closer to everyday colloquialisms than that in the sixteen literary novels (79–80).

Parts of Speech

In the second part of his thesis *Rich Statistical Parsing and Literary Language* (2016), Andreas van Cranenburgh combined readers' ratings with sophisticated NLP methods to investigate which linguistic features correlated more or less with higher or lower perceived literary quality of the 401 novels. He used the Alpino parser for Dutch (Bouma et al.) to tag all 401 novels at the token level for part of speech (noun, verb, article, et cetera) and lemma (the so-called 'dictionary entry'). In Chapter 6, section 1 of his PhD-thesis he explored patterns in the usage of different parts of speech. He pointed out that many handbooks for creative writers advise a sparing use of adjectives and adverbs. Van Cranenburgh showed that novels with high scores for literary quality indeed contained proportionally fewer adjectives and adverbs than lower-rated novels. However, the difference was less striking than expected. Van Cranenburgh also noted a correlation between literary quality and the proportion of nouns and verbs. A normal sentence always has a verb; the relative frequency of verbs is higher in a text with short sentences than in a text with long sentences; there, we find proportionally more nouns. So this could be related to a difference in sentence length.

Van Cranenburgh also examined the proportion of common words, of clauses spoken by characters (dialogue) in relation to descriptive text by the narrator, and the average length of a sentence. For all these properties, he found that they correlated statistically significantly with readers' scores for literary quality. Novels that readers rated highly

literary had longer sentences, relatively fewer common words, more adverbial clauses, and less dialogue. Again, the properties investigated were not independent of each other. For example, more dialogue can also lead to shorter sentences.

Cliché Expressions

Van Cranenburgh also looked into the use of cliché expressions, wanting to check the intuition that platitudes were used more often in novels that were considered to have the least literary quality (section 6.2 of his PhD-thesis and Van Cranenburgh 2018). For this study he was granted permission by Wouter van Wingerden and Pepijn Hendriks to use the collection of cliché expressions that was the basis for their book *Dat hoor je mij niet zeggen!* (**You will not catch me saying that*) published in 2015. They defined clichés as common phrases that people often use without realizing that they are so common. These much-used expressions are sincere reactions to a statement or situation and often function as a linguistic lubricant to keep a conversation going (Van Wingerden and Hendriks 10). I will give an example from the chapter 'Verhaal' ('Story') and present both the Dutch phrase and an English approximation of that phrase:

Ik heb goed nieuws en slecht nieuws. (I have good news and bad news.)
 Zeg jij het of zeg ik het? (Do you say it or do I say it?)
 Ik hóór een 'maar' aankomen... (I hear a 'but' coming...)
 Maar nou komt het! (But here it comes!)
 Dus zo is het gekomen. (So that's how it came about) (Van Wingerden and Hendriks 41)

Van Cranenburgh counted the occurrences of Van Wingerden and Hendrik's list of 5771 cliché phrases in the 401 novels in our corpus. He observed that most clichés were found in dialogue. With an equal amount of dialogue, novels from the category Literary novel indeed contained fewer clichés than those from the categories Suspense and Romance. Most clichés were found in the category Romance, the lowest rated genre in our corpus. A striking exception is J.J. Voskuil's considerably higher rated Literary novel *De buurman* (**The Neighbour*), which contained a lot of clichés. In Chapter 5, I presented a quote from this novel, which occurred as an outlier in my measurements based on word frequencies. I suggested the very high proportion of rather homely dialogue may be the cause.

Novels with hardly any clichés were found in all categories of the corpus, for example E.L. James's *Vijftig tinten grijs* (*Fifty Shades of Grey*), which received the lowest score for literary quality. Very few Literary novels did not have any clichés at

all. Renate Dorrestein's *De leesclub* (**The Reading Group*), discussed in Chapter 4, was one of these. Van Cranenburgh concluded that a high proportion of clichés usually correlated with low scores for literary quality, but that the absence of clichés did not increase the chances of a higher literary rating.

Themes or Topics

To investigate whether highly literary novels paid attention to slightly different themes and topics than novels that were considered less literary, Andreas van Cranenburgh applied topic modelling in section 6.3 of his PhD-thesis. This method involves analysing which words recur in close proximity to each other. Van Cranenburgh generated a set of fifty topics from the parsed corpus. In consultation with his team members he gave each of the topics a label indicating the probable subject or theme. He then calculated the weight of each topic in all novels. He observed that books from the category Literary novel mostly showed a varied set of main topics, whereas books from the genres Romance and Suspense mostly seemed to focus on a single main topic.

One way to explain this discrepancy is the assumption that publishers label a novel as Romance or Suspense precisely because it has one main theme (a troubled relationship, a shocking murder), whereas novels that cover more areas and are therefore more difficult to label are classified as literary novels by elimination (cf. Squires 4–5). Another explanation could be that Literary novels have comparable subjects and themes to novels from the genres of Suspense and Romance, but that the ways in which these are expressed in language are more varied than in the novels from other genres. Either or both of these two situations may apply. An interesting observation is that there is one topic that is characteristic of a small group of originally Dutch-language books in the category Literary novel that also received high scores for literary quality. One of the words in this topic 29, described by Van Cranenburgh as 'music/performance/misc', is *fiets* (bicycle) – which reflects the ubiquity of bicycles in the Netherlands, but is surprising as a possible 'literary' feature.³

All in all, there were few clues as to what exactly were the thematic commonalities of the most respected Literary novels. They all dealt with family relationships (a confirmation of what Archer and Jockers found in *The Bestseller Code*, see Chapter 1), and many of the main characters had an artistic profession, but there was no concrete shared theme or subject.⁴ Van Cranenburgh suggested that this is an

3 The word *fiets* (bicycle) is used in two Dutch novels with cycling as a topic (Smeets, *De afrekening* – **The Reckoning* and Van Sambeek, *Koninginnenrit* – **Queen's Ride*), but the frequency of occurrence in these novels is rather low. Many other novels refer to bikes much more often.

4 More about characters in a corpus of Dutch novels published in 2012 in Van der Deijl et al.

indication that it may be the style of these novels that gave them their high literary scores, rather than the themes of their content.

Predicting Scores for Literary Quality

In order to find out which textual characteristics most closely correlated with a difference in literary appreciation, Van Cranenburgh also applied predictive modelling to investigate the possible relationship between scores for literary quality and the occurrence of combinations of two words (bigrams). In section 7.2 of his PhD-thesis he distinguished between bigrams related to the content of a novel and bigrams that reflect its writing style. An example of a content-related two-word combination is *de trein* (the train). A stylistic example is *terwijl ik* (while I).

It turned out to be fairly easy to predict low scores for literary quality, but it was considerably more difficult to predict which novels were considered highly literary. Van Cranenburgh suspected that this was due to the greater variation of topics in the category of the Literary novel mentioned in the previous section, as opposed to less variation in Romance and Suspense at the bottom of the literary ranking. As an illustration, he presented a short list of the most important content bigrams. Word combinations that stood out for novels that received a high score for literary quality are:

de oorlog, het bos, de winter, de dokter, zo veel, nog altijd, de meisjes, zijn vader, mijn dochter, het boek, de trein, hij hem, naar mij, zegt dat, het land, een sigaret, haar vader, een boek, de winkel, elke keer

(the war, the forest, the winter, the doctor, so much, yet still, the girls, his father, my daughter, the book, the train, he him, at me, says that, the land, a cigarette, her father, a book, the shop, each time)

And for novels which received relatively low scores for literary quality:

de moeder, keek op, mijn hoofd, haar moeder, mijn ogen, ze keek, mobiele telefoon, de moord, even later, nu toe, zag ze, ik voel, mijn man, tot haar, het gebouw, liep naar, we weten, enige wat, en dus, in godsnaam

(the mother, looked up, my head, her mother, my eyes, she looked, mobile telephone, the murder, a while later, (until) now, saw she, I feel, my husband, to her, the building, walked to, we know, only thing that, and so, in god's name)

The bigrams characteristic for less literary text, for example *de moord* (the murder), often suggest suspenseful books. The bigrams that were marked as more literary

were less easy to connect with literary novels – although the bigrams *het boek* (the book) and *een boek* (a book) should be seen as significant here. In the style bigrams, Van Cranenburgh found that the less literary novels contained more word combinations with a question mark (and therefore more questions), and also had more numerals. Those novels also exhibited more informal language usage.

Striking was the occurrence of *mobiele telefoon* (mobile phone) among the bigrams that were mostly associated with lesser literary quality. Van Cranenburgh, who carried out the analysis together with Corina Koolen, noted that even when set in contemporary times, the literary novels used in this experiment apparently did not explicitly mention these gadgets. He suspected that this is because authors of literary novels tried to pursue a certain timelessness (see also Van Cranenburgh and Koolen 2015).

Van Cranenburgh's predictions turned out quite well. However, for several novels the mean score for literary quality was predicted wrongly. For instance, *Kamer (Room)* by Emma Donoghue was assessed as too low and Elizabeth Gilbert's *Eten, bidden, beminnen (Eat, Pray, Love)* as too high. These two novels were discussed at length in Chapter 4, where I concluded that the persistent child's perspective in *Kamer (Room)*, which played a role in the novel being assessed as highly literary, was likely also responsible for a use of language that deviated from other literary novels. For *Eten, bidden, beminnen (Eat, Pray, Love)* I would suggest that readers rated it lower on literary quality because the book is experienced as a 'typical women's book' and for many readers, being labelled a 'women's book' is a swearword for 'non-literary'.

The bigram approach simply dealt with words and punctuation marks as they appeared one after the other and did not take the grammatical position of the words in a sentence into account. Van Cranenburgh investigated whether he could use logical components of grammatical sentences for a similar prediction of reader opinions. A translation of Van Cranenburgh's example sentence from a chick lit novel from the pilot research into syntactic structures, '[h]is jawline is almost square', can be used to explain his approach. He looked at the bigrams 'his jawline', 'almost square', and 'his jawline is almost square' but not, for example, of 'jawline is' or 'is almost', which would be included in the simple bigram approach. In section 7.3 of his PhD-thesis not the tokens themselves were analysed, but the parts-of-speech tags describing the words, for example pronoun + noun, adverb + adjective, and so on. In this experiment, Van Cranenburgh found more sentence fragments that correlated with high literary appreciation than with low literary quality scores. Again, this may have something to do with differences in sentence length: the higher the literary quality, the longer the sentences, and the longer a sentence, the more word combinations can be included in the analysis – but this explanation is not comprehensive. Van Cranenburgh's experiments confirmed that the novels

deemed highly literary contained a larger number of different sentence structures than the novels that were perceived as much less literary.

Finally, in section 7.4 Van Cranenburgh combined the results of various experiments and examined how well they predicted the scores that the participants in The National Reader Survey gave to all 401 novels. In addition to the token bigrams and part of speech bigrams, he used average sentence length, proportion of dialogue, use of the three thousand most frequent words, and the presence of themes (the topics). When he only used these measurable properties of the novels, the model explained 59.7 per cent of the variation in reader opinions. After adding three text external data, namely genre, translation, and author gender, it explained 76.1 per cent.

A model is, by definition, a simplification of reality, and it is not surprising if it does not produce a 100 per cent score. However, Van Cranenburgh's model worked remarkably well. Readers' opinions of most books did not deviate too far from the predictions. And it is fascinating to see which titles were being misjudged. The novel that the model found most deviant compared to the readers is Mart Smeets's *De afrekening* (**The Reckoning*). Van Cranenburgh's results confirm my suspicion that this novel was rated lower by readers than it perhaps deserved on the basis of its writing alone (see Chapter 5 for more about this novel). The same turned out to be true for *Eten, bidden, beminnen* (*Eat, Pray, Love*) by Elizabeth Gilbert. At the other end of the spectrum, some novels were actually rated lower by the model than by the readers. These included Kathryn Stockett's *Een keukenmeidenroman* (*The Help*), Emma Donoghue's *Kamer* (*Room*), Jonathan Franzen's *Vrijheid* (*Freedom*), and the number one literary novel by Julian Barnes, *Alsof het voorbij is* (*The Sense of an Ending*). In an article he wrote together with Rens Bod, Van Cranenburgh added some other novels that the model rated lower than the readers did: Haruki Murakami's *Norwegian Wood* and *1q84*, J.J. Voskuil's *De buurman* (**The Neighbour*), Arnon Grunberg's *Huid en haar* (**Hair and Hide*), Anouschka Voskuil's *Dorp* (**Village*), Niccolò Ammaniti's *Ik en jij* (*Me and You*), and Gerbrand Bakker's *De omweg* (*The Detour*) (Van Cranenburg and Bod 1232). Many of these books already figured as outliers in experiments I described in earlier chapters.

The conclusion Van Cranenburgh drew at the end of his dissertation was that the linguistic characteristics of the novels he analysed contributed to a significant extent to the literary value attributed to the novels by the participants in The National Reader Survey. The score of 59.7 per cent would undoubtedly rise when more literary-sociological aspects were studied, but at least proved robust enough to conclude that the text itself plays an important role. The idea that literary value is solely dependent on extra-textual variables such as literary prizes, author prestige, and so on, could definitively be ruled out.

Vector Space Explorations

Van Cranenburgh emphasized it was not clear whether the relationship between stylistic characteristics and opinions on literary quality can be seen as causal. For that, we would need to observe how readers actually assess the literary value of a text while reading it, but we have not devised a way yet that keeps enough participants on board to yield enough data for analysis. Andreas van Cranenburgh and Corina Koolen (2020) did conduct a pilot study in which they presented a small group of respondents with passages from a number of novels without mentioning the author or the genre, but there are no concrete plans yet to include this approach in a new survey. Van Cranenburgh did come up with the idea to simulate such research, by way of a preliminary step. He wondered whether a model could assess novels at the page-level (Van Cranenburgh et al. 2019, 626–628). To do just that, he divided the novels into passages of about one thousand consecutive words (tokens), always ending at the end of a sentence. For the 401 books in our corpus, which together contain over 52 million tokens, this meant over 52,000 passages of one thousand words long. The whole corpus has a total of over five million sentences. For all those one-thousand-word passages, Van Cranenburgh calculated per novel how big the variation in language use was, and how these related to the reader's ratings of the literary quality of the complete novel. In this way we may see, for instance, that a young adult novel like *Fatale liefde* (*Fatal Love) by Carry Slee has much less variation than a Literary novel like *Caesarion* (*Caesarion*) by Tommy Wieringa. Furthermore, when taking into account the order of the passages, we could see that the alternation between differently styled passages in *Caesarion* is considerably larger than in *Fatale liefde* (628–634).

To identify which themes and which words in the one-thousand-word passages had the highest correlation with literary quality, Van Cranenburgh looked at the fifty Literary novels with the highest and the fifty Literary novels with the lowest mean score for literary quality. To these he applied topic modelling, using the same settings as in his thesis. The fifty novels with the lowest literary rating had a higher proportion of words that can be associated with the following themes:

(Non-)verbal communication

Author: Auel (containing words that can be clearly traced back to the novel *Het lied van de grotten* (*The Land of Painted Caves*) by Jean M. Auel

Settling down

Children

Looks and parties

More prominently represented in the highest-rated literary novels were words around the themes:

Education
 Self-development
 Time, life, and death
 Writers
 Music/performance/misc

The highly rated Literary novels had a larger proportion of abstract themes, while the least valued Literary novels seemed to focus more on social themes and personal relationships. Van Cranenburgh made another interesting observation about the personal pronouns spelling variants *mij/me* (me), *jij/je* (you), *wij/we* (we), and *zij/ze* (both singular she and plural they). The more formal spelling with an -ij occurred significantly more often in the novels that received a high rating for literary quality, and the more informal forms with -e were emphatically more prominent in the literary novels at the bottom of the literary ranking (644–645).

A final experiment used only the 98 novels that received a mean score for literary quality higher than 5 on the 7-point scale. Van Cranenburgh selected passages of one thousand words that were overestimated or underestimated by his model. The underestimated passages – i.e. from novels that readers gave a higher score for literary quality than the model assigned to those passages – turned out to contain topics that can be related to dialogue and violence. The overestimated passages – that is, from novels to which readers gave a lower score for literary quality than the model attributes to those passages – contained topics related to family, religion, and life and death (645).

Comparing the word use of the most overestimated passages with that of the most underestimated passages solidified previous observations. Underestimated passages contained more female personal pronouns, more *ik* (I) and *jij* (you), more words dealing with emotions, more indications of dialogue (*zegt, zei* – says, said), and more negations (*nee, niet* – no, not). Overrated passages consisted of more male personal pronouns and more nouns indicating male roles, *shah*, for example (645–646).

An important outcome of this research is that we can already predict fairly well how the novel as a whole will be judged by readers in terms of literary quality, even for text passages as short as two to three pages. Furthermore, we have a basic grasp of which aspects of the text influence its perception as literary. However, follow-up research is still needed, such as repeat experiments to trace conventions of literariness over time. And a great deal is not yet possible – we could not properly measure plot progression, for example, although currently, a lot of work is being done that may present new opportunities in the future (647–648).

Sentence Length

In all the experiments described above, sentence length seemed to be a text characteristic that correlated remarkably well with readers' judgements on literary quality. In this section, I report on ongoing research into sentence length by Joris van Zundert and myself. For the measurements, Van Zundert made use of the open source software library spaCy (a Dutch language model trained on journals); we measured the average number of words per sentence. Figure 6.1 shows that the higher the literary quality, the longer the average sentence length. However, there was no discernible reverse effect: novels with a long average sentence length are not necessarily experienced as very literary by the readers.

Most of the 401 novels in our research corpus, 322, have an average sentence length of ten to fifteen words. Of the 79 remaining books, 43 novels have an average of fifteen or more words per sentence and 36 have less than an average of ten words per sentence. Of the books with an average sentence length of fifteen or more words per sentence, most (32) were from the category Literary novel and scored 5 or higher for literary quality. But long sentences also occurred in novels from other categories. Take the Suspense novel *Duivelsadem (Arctic Drift)* by Clive Cussler and Dirk Cussler, with a mean sentence length of 18.53 words and a meagre 3.35 for literary quality, and the Romance novel *Vriendschap, liefde en andere stommititeiten (Friends, Lovers and Other Indiscretions)* by Fiona Neill, average sentence length 17.32 words and a score of 3.45 for literary quality. Furthermore, some Literary novels with low scores for literary quality featured rather long sentences, which did not grant them increased literary prestige; this is the case for Elizabeth Gilbert's *Eten, bidden, beminnen (Eat, Pray, Love)*. We also have to keep in mind that in translations, we cannot rule out the influence of the original language and the literary conventions in that culture, as well as the influence of the attention and time translators were able to devote to the translation. Phrasing something concisely usually takes more time than expressing something with many words. These are all aspects that are difficult to investigate.

All categories contained novels with an average sentence length of less than ten words per sentence. The book with the shortest average sentence length was Jens Lapidus's *Snel geld (Easy Money)*, Suspense translated from Swedish. It has an average of 7.84 words per sentence – almost a third of that of the book with the highest mean sentence length, Isabel Allende's Literary novel *Het eiland onder de zee (Island Beneath the Sea)*, translated from Spanish, with 21.07 words per sentence. Only four novels with less than ten words per sentence received a score of 5 or higher for literary quality. One of these was *Kleine dagen (*Little Days)* by Bernard Dewulf, which was awarded the prestigious Libris Literature Prize in 2010. The novel consists of short texts with the author's observations of his two young children, and

the jury described it as ‘a fascinating collection with miniatures about the most ordinary but also the most precious thing: one’s children’ (Cf. <https://www.librisprijs.nl/bernard-dewulf-kleine-dagen/>, my translation.) The short sentences of *Kleine dagen* seem to fit well with the structure and content of the award-winning book.

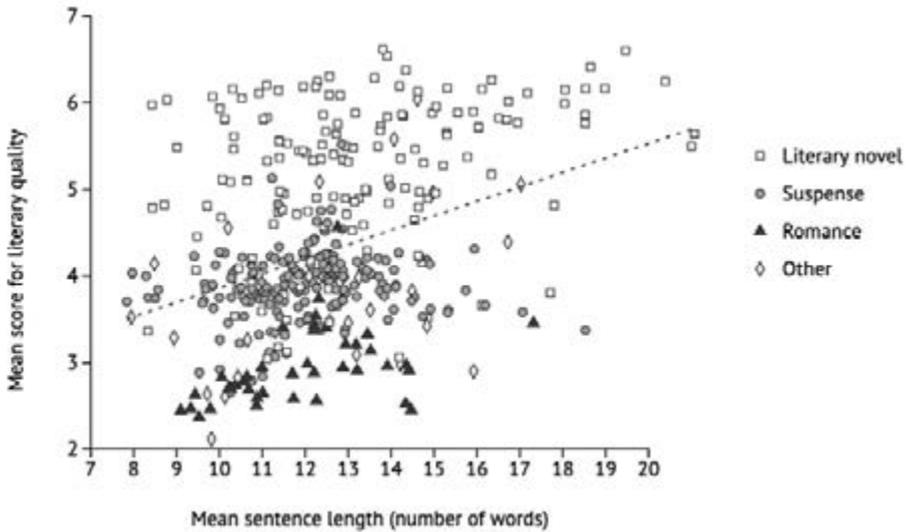


Figure 6.1: Mean sentence length for all 401 books

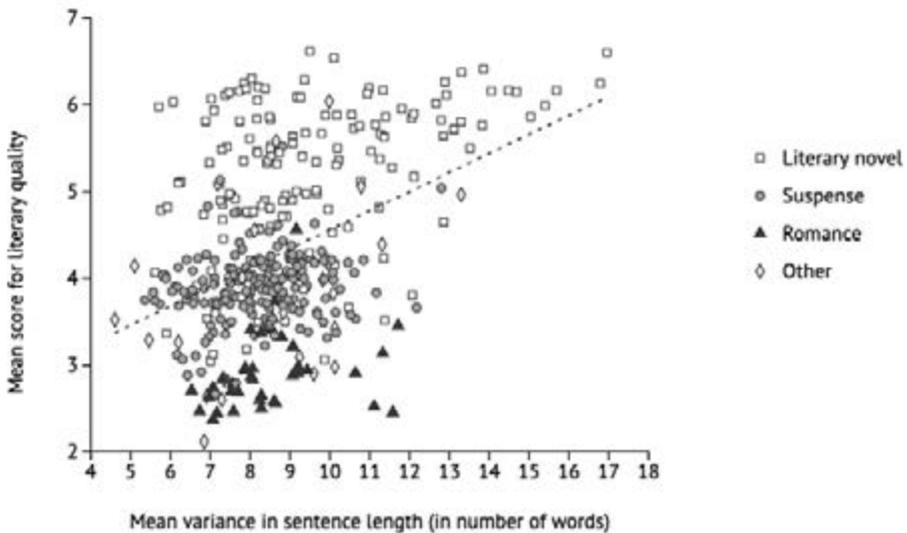


Figure 6.2: Sentence length variation for all 401 books

In Figure 6.2 the variation in sentence length per novel, i.e. the variance, is plotted, showing how large (in number of words) the variation is compared to the average sentence length in the corpus. The book with the highest score for literary quality, *Alsofhet voorbij is* (*The Sense of an Ending*) by Julian Barnes has an average sentence length of 13.8 words, and a sentence length variation of 9.5 words. That variation is the average deviation in word count from the average of 13.8 words per sentence. More details can be found on the companion website or in the litRiddle R package (see Appendix 3).

The ranking of sentence length variation in all 401 books was very similar to that of the average sentence length. There were, however, a number of differences. When sorting sentence length variation from high to low, we found books from the category Literary novel again at the top of the list, while further down the ranks the three main genres were mixed. All in all, average sentence length and sentence length variation appeared to be fairly sound predictors of the literary quality that readers attributed to the novels. In his dissertation Van Cranenburgh convincingly showed that the complexity of the sentences themselves also played a role. However, measuring sentence complexity is not as easy as measuring sentence length and its variation. To my mind, a measurement of the average sentence length can therefore function very well as an indicator for literary quality and as a starting point for further exploration.

Comparing Dutch Literary Novels by Female and Male Authors

In the two previous chapters Literary novels written by female and male authors were discussed separately, because readers clearly rated them differently. In this section I bring them together again, limiting my final experiment to Dutch-language originals only. Table 6.1 juxtaposes the 10 most highly rated novels of both female and male authors and their ratings so that the differences between the mean scores for literary quality can be easily verified. The top 10 novels written by men were virtually at the top of the total ranking, the first one in position 2 and the tenth one in position 16. The parallel top 10 of books by women were spread much more widely over the ranking: from position 36 to 103 – much lower. The difference between the average score for literary quality for the two numbers 1 is 0.7 points, which increases to 1.23 points for the two numbers 10. This difference is visualized in Figure 6.3.

Table 6.1: The ten novels by Dutch-language female and male authors with the highest scores for literary quality

Rank	Author is male	Title	Score M	Score F	Author is female	Title	Rank
2	Erwin Mortier	<i>Godenslaap</i>	6,60	5,90	Margriet de Moor	<i>De schilder en het meisje</i>	36
3	Erwin Mortier	<i>Gestameld liedboek</i>	6,54	5,86	Marente de Moor	<i>De Nederlandse maagd</i>	42
5	Tom Lanoye	<i>Sprakeloos</i>	6,37	5,81	Anna Enquist	<i>De verdovers</i>	53
7	A.F.Th. van der Heijden	<i>Tonio</i>	6,29	5,66	Jessica Durlacher	<i>De held</i>	56
8	Stephan Enter	<i>Grip</i>	6,26	5,53	Renate Dorrestein	<i>De stiefmoeder</i>	58
9	J. Bernlef	<i>Geleende levens</i>	6,25	5,36	Franca Treur	<i>Dorsvloer vol confetti</i>	64
11	Tommy Wieringa	<i>Caesarion</i>	6,20	5,35	Vonne van der Meer	<i>De vrouw met de sleutel</i>	81
13	Gerbrand Bakker	<i>De omweg</i>	6,19	5,11	Joshua Zwaan	<i>Parnassia</i>	91
15	Arthur Japin	<i>De overgave</i>	6,17	4,98	Renate Dorrestein	<i>De leesclub</i>	99
16	Thomas Rosenboom	<i>Zoete mond</i>	6,17	4,96	Susan Smit	<i>Vloed</i>	103

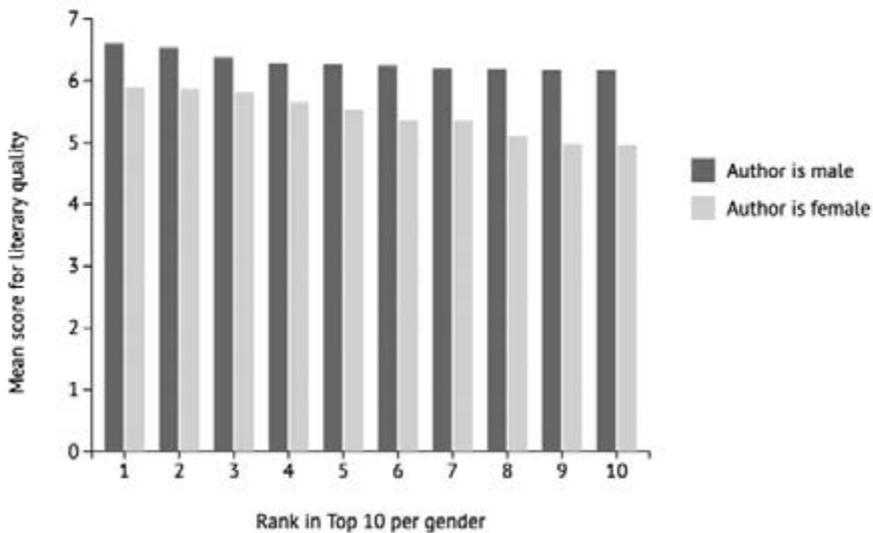


Figure 6.3: Literary quality top 10 of Dutch-language novels by female and male authors

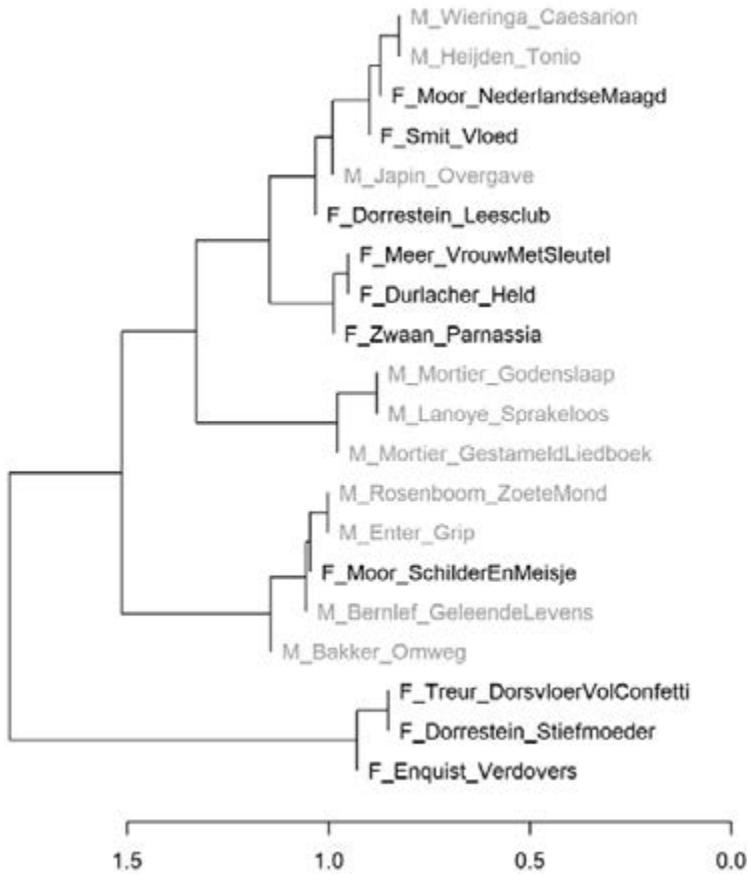


Figure 6.4: Literary quality top 10 of Dutch-language novels by female and male authors, Cluster Analysis (1000 most frequent words)

A cluster analysis of the vocabulary and word frequencies of the thousand most frequently used words (Figure 6.4) showed that the literary novels written by women and men do not differ significantly in the use and frequency of those words. A PCA (Figure 6.5) did not yield a sharp division either; the novels written by women and men did not fall into two mutually exclusive groups. I used the Stylo package for R (Eder et al.) as I did in earlier chapters; more information about the methods can be found there. Both measurements had a couple of outliers compared to the rest. From the male-authored novels these are the three books by Mortier and Lanoye, whose Flemish background may partly explain this (Figure 6.5 at the top). *De omweg* (*The Detour*) by Gerbrand Bakker is located farthest away from the other nineteen books (Figure 6.5 bottom right). It is not yet clear to me why this novel is an outlier in this selection of novels. From the female-authored

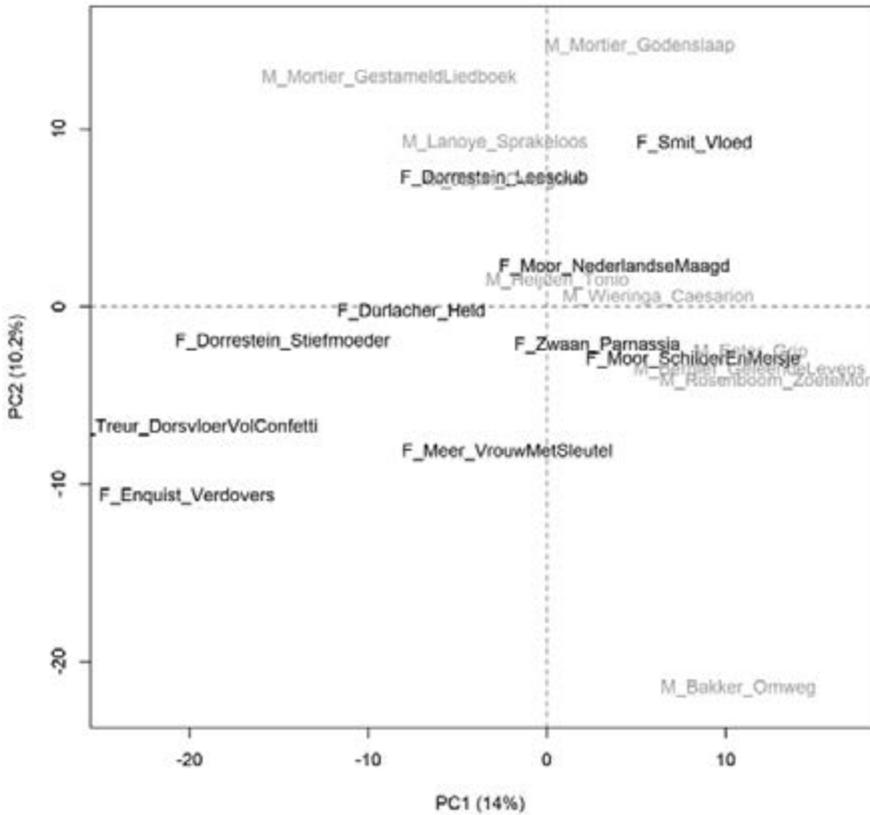


Figure 6.5: Literary quality top 10 of Dutch-language novels by female and male authors, PCA (1000 most frequent words)

books, Figure 6.5 shows three novels that are clustered a bit to the side on the left: *Dorsvloer vol confetti* (*Confetti on the Threshing Floor*) by Franca Treur, *De stiefmoeder* (*The Stepmother*) by Renate Dorrestein, and *De verdovers* (*The Anaesthetists*) by Anna Enquist. The reason for this clustering in the margin of the rest has not yet been clarified either. For now, the most important observation is that the measurements do not show a clear separation between the novels written by female and by male authors and do not help to textually explain the gendered gap in ratings.

In the previous section I discussed sentence length in more detail. Figure 6.6 shows the mean sentence length and the mean score for literary quality for all fifty titles by male authors and all 23 by female authors – these include the eight titles published separately as stories or collections of stories, see chapter 4 and 5 for more details. Note that the outliers on the side of high mean sentence length were all novels by male authors; any novels with that kind of sentence lengths written

by women did not make it to the list of 401 most sold and most borrowed novels which ended up in our corpus.⁵

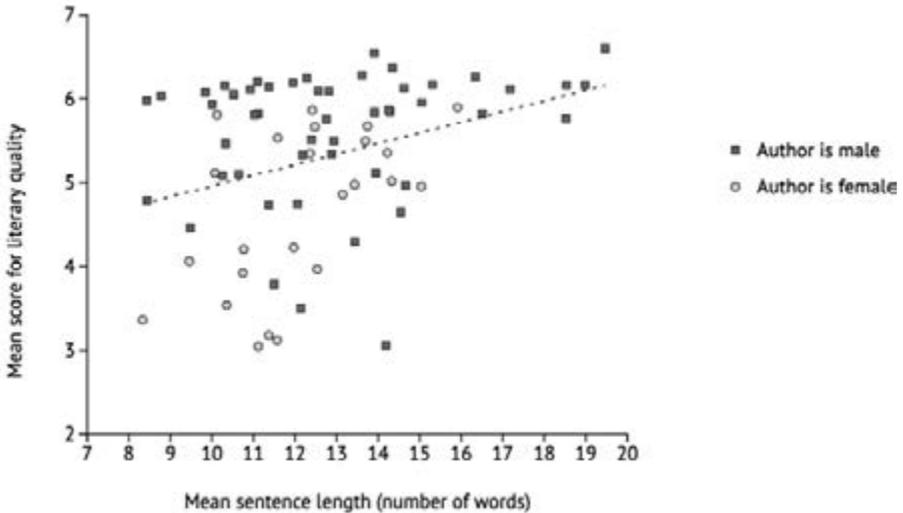


Figure 6.6: Mean sentence length (in words) of Dutch-language Literary novels

All in all, in terms of sentence length, the books written by men and women do not deviate significantly from each other. The same holds for the variation in sentence length, the average deviation in number of words from the average number of words per sentence (graph not shown; an interactive version of it and of Figure 6.6 shows author names and titles of the novels, see Appendix 3). Sentence length here does *not* function as a reliable indicator for reader ratings of literary quality.

In Chapter 4 and Chapter 5 I compared the vocabulary of the Literary novels that received the highest literary appreciation from the readers with that of the Literary novels that received the least appreciation using the keyword function in Laurence Anthony's AntConc tool (2019). These comparisons showed two clear trends: differences in the use of personal pronouns and in the use of present or past tense verb forms. These results are placed next to each other for easier comparison in Table 6.2 and Table 6.3.

⁵ In Figure 6.6, the seven novels by male authors (represented by the small grey squares in the plot right) that have longer sentences than the first novel by a female author (the lighter circles in the figure) in the literary rankings occur at place 2, 8, 16, 17, 23, 46, and 51.

Table 6.2: Personal pronouns more prominent in novels with the highest scores for literary quality compared to those with the lowest scores and vice versa

Literary novels: personal pronouns	More prominent in novels with highest scores for literary quality	More prominent in novels with lowest scores for literary quality
Female authors, translated	<i>ik, mijn, me</i> (I, mine, me)	<i>je, jullie, ze, haar</i> (you (sing.), you (plur.), she/they, her)
Female authors, Dutch-language originals	<i>hij</i> (he)	<i>ik, me, mij, mijn, mezelf, jij, je, jou, jouw, we</i> (I, me, me, mine, myself, you (sing.), you (sing.), you (sing.), your (sing.), we)
Male authors, translated	<i>ik, je, mij</i> (I, you (sing.), me)	<i>hij, hem, ze, hen</i> (he, him, she/they, them/their)
Male authors, Dutch-language originals	<i>me, ge, hij, hem, haar, zij</i> (me, you, he, him, her, she/they)	<i>we, jullie</i> (we, you (plur.))

Table 6.3: Verb tense more prominent in novels with the highest scores for literary quality compared to those with the lowest scores and vice versa

Literary novels: verb tense	More prominent in novels with highest scores for literary quality	More prominent in novels with lowest scores for literary quality
Female authors, translated	Present tense	Past tense
Female authors, Dutch-language originals	Past tense	Present tense
Male authors, translated	<i>is, heb</i> (is, have)*	Past tense
Male authors, Dutch-language originals	Present tense	Past tense

* In the top 100 of words only these two verb forms occur, mainly functioning as auxiliary verbs or copula. This evidence is too sparse to state here that the present tense is more prominent than the past tense.

As shown in Table 6.2, the trend in personal pronoun use is slightly different for each subgroup. The most striking observation is that the Dutch-language novels written by female authors and valued highest for their literary quality more often contained the third-person pronoun singular *hij* (he), and the least literary novels in that group more often contained first-person singular or second-person singular pronouns. This differs from what we see in the other three groups, where we find more first-person singular pronouns in the novels that received the highest literary appreciation, while in the least literary group other pronouns are more prominent. The most literary Dutch-language novels written by women therefore seem to refer more frequently to male characters in a story told in the third person, whereas the Literary novels to which less literary quality is attributed more often refer to

a first-person singular or plural. The strikingly more frequent occurrence here of second-person singular pronouns could indicate a higher amount of dialogue. Translated novels that are considered highly literary and written by men show an increased presence of first-person pronouns, whereas the least literary novels favour a male third person. In the highest-scoring Dutch-language Literary novels written by men, we find great diversity in references to persons, whereas the books with the lowest scores for literary quality tend to utilize *we* and *jullie* (we and you, plural) relatively more often. As such, they seem to express more sense of community versus more personal reflection.

In Table 6.3, in which the predominant verb tenses are listed by group, we see again that the novels by Dutch-language female authors stand apart from the other groups. The past tense is more prominent in the novels that receive little literary appreciation from readers in three of the four groups, with the exception of the books by Dutch-language women, where the present tense predominates at the lower end of literary quality ratings. This group was the only one to have a more prominent use of the past tense in its most highly literary novels.

The observed differences seem to be related to two main characteristics of the novels: first, the narrative perspective, with usually either first-person or third-person narration, or a mix. Second, the proportion of dialogue. Because dialogue text can be expected to resemble first-person narration rather than third-person narration, first-person narrative may be difficult to distinguish from dialogue automatically. Indeed, although methods have been developed, there are no computational methods yet which can adequately measure for a large amount of novels which perspective or which changes of perspective each novel has. Neither is it possible yet to automatically distinguish dialogue from narrative with acceptable results. Manually annotating even a small corpus of novels is labour-intensive and our complete corpus of 401 novels is way too large for this. We did make a start with the development of methods that should make larger-scale research into perspective and dialogue possible (Van Rossum et al.). However, the corpus we used for our first attempts also was too small to draw any firm conclusions based on the observed trends described above. They simply ask for more research.

I will end this section with a keyword analysis of the top 10 most literary novels by female and male Dutch-language authors, again performed with Anthony's AntConc software, to establish which words occur significantly more frequently in novels written by female authors compared to those written by male writers and vice versa. Although again, the number of novels analysed is way too small for firm conclusions, we do find an interesting difference in the occurrence of nouns that indicate family members. In the ten novels written by women, *papa*, *vader*, *mama*, *oma*, *opa*, *stiefmoeder*, and *moeder* (dad, father, mum, grandma, grandpa, step-mum, and mother) occur significantly more often than in the novels authored by men; in

the latter, *schoonmoeder*, *schoonvader*, and *oom* (mother-in-law, father-in-law, and uncle) occur more frequently than may be statistically expected. For these novels, we can deduce that in this small subset, the books written by women probably devote more attention to family relationships than those written by men, or at least that they deal with a greater variety of family relationships.

Author Gender and Perceptions of Literary Quality

Whether and how the author's gender played a role when readers formed an opinion about a novel has been investigated by Corina Koolen as part of the project *The Riddle of Literary Quality* in her PhD thesis *Reading Beyond the Female: The Relationship Between Perception of Author Gender and Literary Quality*. I will summarize the main outcomes of her research here; for more details the reader is referred to Koolen's PhD-thesis.

To put the observed differences in perceived literary quality of novels written by female and male authors in perspective, Koolen investigated the gender balance in the Dutch literary field to verify whether it perhaps was numerically dominated by male authors. This appeared not to be the case. Koolen found that in the years 2007–2012, approximately as many women as men were active as professional writers in the Netherlands. This means that women made up about 50 per cent of all publishing authors. The share of women dropped to between 35 and 40 per cent for all original Dutch-language books published with the label 'literary novel' (more about the labels in Chapter 3). The attention of critics was even lower: of all reviews devoted to literary novels, only 30 per cent addressed works by female authors. Literary prizes – which most unequivocally signal literary acclaim – were awarded to novels by female authors in only 21 to 25 per cent. From a longitudinal perspective, the percentages on the upper rungs of this 'literary ladder', as Koolen called it, even seem to be going down rather than up. Her conclusion, therefore, was that the literary standing of female authors in the Netherlands was not equal to that of male authors, and that there was no sign of an upward trend (Koolen 39–44). Koolen also debunked the optimism that a breakthrough for female writers was impending by demonstrating that similar discourse has prevailed for centuries, while few female-authored texts attain longevity into the present day (52–56).

In our research corpus of 401 novels, female and male authors were represented in roughly equal proportions. The distribution of economic capital along gender therefore roughly corresponded to the 50–50 percentage of women and men who are professionally active as writers. Of the original Dutch-language Literary novels in the corpus, 33.3 per cent were written by a woman, which is slightly lower than the 35 to 40 per cent Koolen found for the Dutch book market (65).

In the Suspense category, books written by women and men were rated equally in terms of literary quality by the participants in The National Reader Survey. This was not the case for the Literary novel. Koolen wondered whether the gender of the survey participants played a role. Leaving aside the 0.6 per cent (ninety people) of the 13,784 participants who did not share their gender, the male respondents read mainly books written by men. The women read books by both male and female authors, with men being represented slightly more. When it came to appreciation, men gave slightly higher scores to female authors than women did. But there were far fewer male participants. On the whole, female authors received lower average scores than male ones (70–77).

For Suspense, Koolen noted that female readers rated novels by female and male authors equally on literary quality, and that male readers rated novels by female authors slightly higher than those by male authors. For the Literary novel, there was a difference in appreciation among both female and male readers: novels written by male authors were rated more highly. For female readers, the difference was slightly larger than for male readers (77–83).

To see whether this difference was also reflected in the motivations readers gave for one of the books they reviewed, Koolen compared the vocabulary in different sets of open responses. She noted (as I did in chapter 2) that women and men both took the genre of the novel in question into consideration. And zooming in on the motivations of highly literary novels, she found that female readers phrased their opinions about male novels slightly more often in literary-technical terms that referred to good style, structure, and deeper layers. Writing about female novels, they referred more frequently to the content and the story (83–95).

To find out whether the knowledge of an author's gender alone influences the reader's assessment of the literary quality of a text, Koolen developed a new experiment. She asked literary author Emy Koopman to write a short text with gender-neutral content. Koolen then presented this text to six hundred test subjects as a manuscript of a novel that had been submitted to a publisher. It was presented as the author's first novel and the test subjects were asked how literary they thought the text was based on the short sample. Of the six hundred test subjects, two hundred panel members read in the introduction that the author was male, two hundred that the author was female, and in the introduction for the third group of two hundred panel members, the author's gender was not mentioned. Koolen expected that readers who were under the impression that the author was female would attribute less literary quality to the text than those who thought the author was male, and that female panel members would be more critical than male subjects. To her surprise, both expectations were disproven (96–101). As possible explanation, Koolen suggested that female and male authors roughly had the same literary potential in the eyes of the respondents because of the explicit mention that the

text passage was from the author's debut. It could therefore be the case that it was not exclusively the author's gender that played an important role in readers' appreciation, but another aspect, namely the prestige, the reputation of (already well-known) authors (101–103). In 2013, these were still more often men than women, and that numerical overview may have influenced readers.

Even then, the question remained how readers were consciously or unconsciously influenced by the gender of the author. To explore whether the 'feminine' or 'masculine' nature of a novel may have played a role, Koolen examined how 'feminine' a novel was perceived on three levels: genre, style, and subject/theme. Regarding genre, Koolen showed how words like *vrouwenboek*, *damesroman* (women's book, ladies' novel), and so on, as used exuberantly by readers in their substantiations of ratings of literary quality, had an overwhelmingly negative connotation. Koolen found a total of 369 such qualifications in the motivations readers presented. Koolen also pointed to the possibility that novels written by women were, by definition, marked by readers (and publishers) as belonging to a genre of 'women's books', resulting in points deducted for literary appreciation. Cause and effect (is it primarily about the genre or about the author's gender?) were thus not always clear. In their motivations, readers never explicitly mentioned the gender of the (female) author, but they did mention that certain books were written for a female audience *and* had an accessible style. As far as style was concerned, in the motivations Koolen saw that readers mainly responded negatively to a novel's emotional valence. A certain detachment clearly was considered more literary and emotionality was linked to 'women's books' (114–121).

Readers also associated certain topics and storylines with the feminine or masculine nature of a novel. If it was about love, that lowered the expected literary quality but if and only if the novel was written by a woman. When authors wrote about children, it usually worked against their literary prestige, especially for female writers. Strictly speaking, therefore, while readers may place subject matter in a gendered realm, their appreciation can ultimately not be seen in isolation from the author's gender. Koolen concluded that readers did use references to the *femininity* of subject, style, or genre to explain what was lacking in the literary quality of a novel, but did not reference *masculinity* in the same way (122–126).

For the measurable aspects of style, Koolen emphasized that it is methodologically problematic if research emphasizes only the differences between the writing style of female and male authors. It should not be forgotten that there are many more similarities than there are differences. It must be ensured that a corpus is carefully composed in order to misconstrue author gender for other confounding factors. Because our research corpus of 401 novels was unbalanced in terms of genre, Koolen collected Dutch-language novels from the same period (2007–2012) that had been nominated for a major literary prize. Her research corpus contained 25

novels written by men, 24 by women and one by a transgender man who identified as female when the novel was published. This was the largest proportionate corpus possible because of the limited number of nominated novels written by women. Two novels written by women and three by men actually received a prize (128–136).

Together with Van Cranenburgh, Koolen analysed this corpus of award nominees. She pointed out the difficulties of comparisons between female and male-authored novels, highlighting all kinds of biases in for example the corpus selection. Taking as many of these issues into account as possible, she concluded that originally Dutch-language literary novels written by women could not be clearly distinguished from those written by men. There were some differences she did observe, and they had also been established in other research that she referred to: women used personal pronouns slightly more often and men slightly more prepositions, articles, and numerals. It is not evident how this could be relevant for the analysis of literary quality, however (133, 157–158). Koolen also applied a method called LIWC – Linguistic Inquiry and Word Count. Developed by psychologist James Pennebaker and his team initially for non-fiction, LIWC contains a number of manually compiled word lists (also for Dutch) that are used to obtain an indication of, for example, the proportion of positive and negative emotions expressed in a text, among a number of different categories. In the corpus of literary nominees, the differences between novels written by women and by men for the aspects that can be analysed with LIWC appeared to be minimal, with only one exception. On average, women and men used words related to the aspect ‘body’ about equally often, but the occurrences were unevenly distributed. In the nominee corpus, they occurred in only a small number of novels written by female authors and in those novels they were very frequent, whereas they did occur in lower frequencies in many more novels by male authors. This was surprising, as preoccupation with the body is generally thought to be a typically female subject (Koolen 2018, 136–143).

This prompted Koolen to investigate the subject further. She used the corpus of 32 novels that the team also used for a first experiment into syntactical analysis, which I described at the start of this chapter. In these 32 novels Koolen observed that the proportion of sentences containing a physical feature seemed to be higher in sixteen literary novels than in the sixteen chick lit novels (162–192). To verify this similarly unexpected finding, Koolen did a close reading of the first five hundred sentences of the novels to see whether attention to physical characteristics was indeed a typically ‘female’ subject. She distinguished between four possible configurations: the reader looks through the eyes of a male character at a female or a male protagonist, or through the eyes of a female character at a female or male protagonist. For each of these configurations Koolen examined what the person being looked at looks like, and with what feelings the person is being looked at

(192–204). She also established in which category of novel a configuration mainly occurred, taking into account the gender of the author (204–241).

Koolen summarized that in the 32 novels she used for this analysis, female authors usually (but not always) write about female protagonists and male authors usually (but not always) about male protagonists. The ‘femininity’ of the subject ‘physical characteristics’ involves that it is usually women who are being looked at. In the sixteen literary novels they are usually looked at by men, in chick lit novels they scrutinize themselves. Men are mostly looked at in the sixteen chick lit novels, and then only by women. In the chick lit novels, women are looked at as much as men (242–244). Koolen concluded that the selected chick lit novels and literary novels had more in common as far as the description of physical characteristics was concerned than might be expected. Novels in which male characters viewed what Koolen described as ‘the destructive nymph’ (204) were experienced as literary. Those in which female protagonists viewed ‘the caring Adonis’ (215) were not considered to have literary quality. This confirmed the observation that in the Dutch literary field publishers, reviewers and, moreover, the (predominantly female) readers seemed to perpetuate the convention that novels with female protagonists get attributed less literary value (246). This is closely related to the differing literary prestige of different genres; genre cannot be seen as completely separate from a gender perspective (248).

The Feedback Loop of Genre

Taking all results of the project into account, my conclusion is that the biggest differences between the genres examined in *The Riddle of Literary Quality* and within the category of the Literary novel had to do with forms of complexity in the language used. Romance and the Literary novel were furthest apart in both literary appreciation and style. Suspense was in between. Most books in the category of the Literary novel on average had a larger vocabulary, a longer average sentence length, a larger variety of sentence lengths, and a higher proportion of complex, compound sentences than Suspense novels. Members of the Romance genre showed the opposite: compared to Suspense, they tended to have a smaller vocabulary, shorter average sentence length, less variety in sentence length, and a lower proportion of complex sentences.

There were clear trends, but also exceptions. Sometimes, I could confirm computationally that books with a certain genre label stylistically resembled novels with a different genre label. Readers, however, often did not seem to notice this and attributed a rating for literary quality to the novel that agreed more with the genre label the publisher selected, so with how the novel was marketed, than with

the style of the text itself. What we can conclude from this is that the genre label selected by the publisher had a disproportionate influence on how readers value a novel. And as far as we can see, this influence usually pushed readers' opinions in only one direction: downwards. If a novel was given a 'less literary' genre label, the chances of a high literary rating for the novel were gone. This seems to be the case for the Dutch translation of Candice Bushnell's *1 Fifth Avenue* (*One Fifth Avenue*), which was labelled as Romance but used language that was more in keeping with the Literary novel. If a book had been 'upgraded', such as the *Vijftig tinten* (*Fifty Shades*) trilogy, which the publisher labelled as Literary novels despite its stylistically divergent language use, this did not guarantee a higher rating for its literary quality.

My assumption is that publishers cannot afford to 'downgrade' a novel's genre too often, otherwise they would lose their credibility and reliability as literary gatekeepers. Labelling a novel under a less literary genre would usually provide the publisher with higher sales – in Bourdieu's terminology: economic capital – which secures turnover and therefore income for the publisher. This income then enables new books to appear and find their way to eager readers. It is a different matter, however, for the authors of those novels. Admittedly, authors will usually welcome the economic capital, but not always at the cost of symbolic capital. If they themselves especially aimed to write a Literary novel that was then presented by the publisher as being Suspense or Romance, they will lose their chance at literary prestige – except in the unlikely event that the novel is awarded a literary prize. Being awarded a prestigious prize can seriously lift the readers' appreciation, as demonstrated by the case of Julian Barnes's *Als of het voorbij is* (*The Sense of an Ending*), which received a higher average score for literary quality from the participants in The National Reader Survey than the measurable text properties would have suggested. But Booker Prize winner Barnes was published under the label Literary novel. What would happen if a novel without that label received a literary prize? In the UK in 2018, the nomination of suspense novel *Snap* by Belinda Bauer on the longlist for the Man Booker Prize alone led to major protests. The book did not make it to the shortlist (Taylor 129).

How important genre is in the publishing marketplace was elaborately described by Claire Squires in *Marketing Literature: The Making of Contemporary Writing in Britain*. This book presents a comprehensive view on social, economic, and cultural contexts of the production of writing in Britain at the end of the twentieth and the beginning of the twenty-first century. Squires' main focus was on literary fiction, but she did not exclude genre fiction from her analysis (14). From a marketing perspective, literary fiction can be called a genre, she stated (5); interestingly, in his machine learning approach of genres in a corpus of Spanish novels, José Calvo Tello comes to the same conclusion (Calvo Tello Ch. 5.3). According to Squires, genre 'is a crucial component in the marketplace, as it is one of the primary means by

which authors and readers communicate' (70). Both readers and authors influence the ways genres are styled (74, 84–85).

The author's perspective has not been addressed yet in this book. One of the scholars who theorized the role of genre in the writing process is literary scholar Anis Bawarshi in his insightful 2003 book *Genre and the Invention of the Writer*. Bawarshi described the symbiotic interdependence between genres and writers in which genres are maintained by the writers, to whom they provide space in turn. According to Bawarshi, genres are therefore both the conditions for the creation of new texts and the consequences of the conventions that are propagated in those new texts. It is thus a process of mutual influence. For an author, genres are something they can focus on in the writing process: certain characteristics 'belong' to a certain genre. The writer has the option of conforming to a genre and using those typical characteristics, but also has the freedom to be creative with it. In the latter case, new genres may emerge from existing ones (9). The way of dealing with genre characteristics is thus a dynamic and creative process, and an author has the freedom of choice to occupy different positions in that social process, as Bawarshi emphasized (10–11, 43–45). Not all those positions will be filled by a writer to the same extent. Depending on one's writing experience, education, and/or personal preferences, an author may feel more comfortable with one interpretation than another (99).

Bawarshi also saw a mutual dependence between genre and readers. He referred to the role of university research and education. The fact that 'genre fiction' such as Romance and Suspense is excluded from research and teaching of literature sends a message to students that all works that are labelled not as literature but as genre fiction are not worth studying. This influences readers in what they consider good and bad. The genre label thus makes it clear to readers what is expected of them in their assessment of the works. It is very difficult to escape this process, however much one might wish to do so (Bawarshi 2002, 343–348).

Neither Squires nor Bawarshi indicated *how often* publishers opt for changes in genre labels. This is precisely where computational genre scholarship may play a role. Scholars working in computational literary studies are advocating a bottom-up approach to mutually distinguish different genres. They apply computational methods to large digital collections to see which clusters emerge, and only afterwards look at formal genre labels (for an overview see Primorac et al., and for an example of finding 'hidden subgenres' in this way see Calvo Tello Ch. 6.2). I expect that comparing the results of such measurements with the actual labels attributed by publishers may precisely show where tension arises between marketing strategy and textual form and content. My discussion of the five novels by Henning Mankell in our research corpus in Chapter 3 fits this approach. Three of these novels were labelled as Suspense and two as Literary novels. The two literary novels differed measurably in style from one another, which may have had something to do with

one of them being a recent translation from a much older original published in 1982. The language of the translation of the more recent Literary novel, however, proved to be very similar to Mankell's three Suspense books. Furthermore, the readers' scores for literary quality did not differ much for all five novels. This is unexpected from the perspective of the clear genre hierarchy that was found and makes it difficult to take readers' opinions of Mankell's novels at face value.

In his seminal study about digital analysis of literary genre using machine learning, José Calvo Tello showed that for his corpus of novels from the Spanish Silver Age, '*literary fiction* has a stronger correlation with the literary perception of the author than with that of the text as such' (Calvo Tello Ch. 5.3.4). From this point of view, if readers were aware of Mankell's activity in both genres, this knowledge could perhaps raise their ratings for the literary quality of his Suspense novels. If readers knew him mostly as an author of Suspense, they may have downgraded his Literary novels. In Bawarshi's words, Mankell's name would be linked to suspense as a 'preferred genre', and readers might have based their opinion of his literary novels on that. Both of these scenarios are challenging to test empirically.

Conclusion

This chapter started with a presentation of the results of computational text analysis experiments on the texts in the corpus of 401 Dutch novels that went beyond word frequencies, mostly performed by Andreas van Cranenburgh. It demonstrated that novels that readers perceived as high in literary quality generally had a more diverse vocabulary, more difficult words, longer sentences, and more complex sentences, than those with low literary quality. The most highly valued novels also contained proportionally fewer adverbs and adjectives, relatively less verbs, and more nouns. The ratio of nouns and verbs could be a side effect of the differences in sentence length and sentence complexity. Furthermore, fewer cliché expressions occurred in the most literary novels, although exceptions apply. Topic modelling showed that, compared to Romance and Suspense, Literary novels usually had more different topics, which supported the idea that from a marketing perspective, literary fiction may be seen as a category that cannot be labelled as belonging to existing kinds of genre fiction, so as a residual category.

Predictive modelling based on the texts themselves explained the mean score for literary quality given by the readers with a percentage of 59.7. When some extra-textual metadata were added (genre, translation, gender of the author) this went up to 76.1 per cent. Incorrect predictions in some cases uncovered a probable bias of the readers – for example for Elizabeth Gilbert's *Eten, bidden, beminnen* (*Eat, Pray, Love*) and for Mart Smeets' *De afrekening* (**The Reckoning*) – and in other cases

reflected specific stylistic characteristics of a novel, as for Emma Donoghue's *Kamer (Room)*. I also described the method Van Cranenburgh developed to predict scores based on passages of 2 to 3 pages of text, which may be very useful for follow-up research.

In the second section of the chapter, I brought the Literary novels by female and male Dutch-language authors together again for a final comparison of both ratings and texts. For the complete corpus of 401 novels, the most significant correlation with high literary quality was sentence length. The higher the mean score for literary quality, the longer the mean sentence length. This seemed to apply at least to Literary novels written by male authors, which repeatedly showed several outliers with very long mean sentences. Among the literary novels by Dutch-language female authors there were none with equally long sentences. This does not mean that female authors never used such long sentences. We can only conclude that such novels did not make it to the bestseller list or the top of the list of books most borrowed from public libraries that we used to select the novels for our corpus of 401 books. Broadening the research by adding less well-known novels from the same time period would be a very interesting next step, because it could help to find out if novels by female authors with comparable long sentences to those by male authors would also receive lower scores for literary quality than the novels that did make it into our corpus.

An overview of the results of Corina Koolen's research into the influence of author gender on the perceptions of literariness confirmed that differences in ratings were significant and certainly reflected a bias in the readers' appraisal of literary quality against novels by female authors. Koolen's experiments suggested that this gender bias is not due to a pervasive idea in readers' minds of women not being able to write well. Instead, the bias seemed to reflect the fact that male authors were much more visible as having literary prestige than female authors. Furthermore, it seemed plausible that thematic characteristics of novels that readers generally perceived as feminine led to lower ratings of female-authored novels.

I closed the chapter with a short section that addressed the feedback loop of genre. Publishers and readers hold specific ideas that influence how genres are marketed and perceived. Authors also play a role in this system. They are typically aware of the genre context in which they want to have their work published, and let their writing be influenced by what that genre seems to expect, even when they want to push at the boundaries of genre and through that challenge influence how future readers and publishers perceive it. This again emphasizes how several interconnected factors are at play and how difficult it is to pinpoint what exactly makes a novel highly literary or not. In the next concluding chapter, I will reflect on the main outcomes presented in this monograph and consider the implications and possible next steps for this line of research.

References

All hyperlinks were last accessed in February 2023.

- Anthony, Laurence. *AntConc*. Version 3.5.8., Tokyo: Waseda University, 2019, <https://www.laurenceanthony.net/software.html>.
- Bawarshi, Anis S. 'The Genre Function.' *College English*, vol. 62, no. 3, 2002, pp. 335–360, doi:10.2307/378935.
- Bawarshi, Anis S. *Genre and The Invention of the Writer: Reconsidering the Place of Invention in Composition*. Utah State University Press, 2003.
- Bouma, Gosse, et al. 'Alpino: Wide-Coverage Computational Analysis of Dutch.' *Language and Computers*, vol. 37, no. 1, 2002, pp. 45–59, <https://www.let.rug.nl/vannoord/papers/alpino.pdf>.
- Calvo Tello, José. *The Novel in the Spanish Silver Age. A Digital Analysis of Genre Using Machine Learning*. Bielefeld University Press, 2021, doi:10.14361/9783839459256.
- Eder, Maciej, et al. 'Stylometry with R: A Package for Computational Text Analysis.' *The R Journal*, vol. 8, no. 1, 2016, pp. 107–121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.
- Jautze, Kim, et al. 'From High Heels to Weed Attics: A Syntactic Investigation of Chick Lit and Literature.' *Proceedings of the Second Workshop on Computational Linguistics for Literature*, Atlanta, 14 June 2013, <https://aclanthology.org/W13-1410/>.
- Koolen, Corina. *Reading Beyond the Female: The Relationship Between Perception of Author Gender and Literary Quality*. 2018. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37fice7>.
- Primorac, Antonija, et al. 'Distant Reading Two Decades On: Reflections on the Digital Turn in the Study of Literature.' To be published in *Digital Studies / Le champ numérique*, doi:10.16995/dscn.8855.
- Squires, Claire. *Marketing Literature: The Making of Contemporary Writing in Britain*. Palgrave Macmillan, 2009.
- Taylor, Helen. *Why Women Read Fiction: The Stories of Our Lives*. Oxford University Press, 2019.
- Van Cranenburgh, Andreas, and Corina Koolen. 'Identifying Literary Novels with Bigrams.' *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, Denver, June 2015, pp. 58–67, doi: 10.3115/v1/W15-0707.
- Van Cranenburgh, Andreas. *Rich Statistical Parsing and Literary Language*. 2016. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37fice7>.
- Van Cranenburgh, Andreas, and Rens Bod. 'A Data-Oriented Model of Literary Language.' *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Valencia, 3–7 April 2017, pp. 1228–1238, <https://aclanthology.org/E17-1115/>.

- Van Cranenburgh, Andreas. 'Cliché Expressions in Literary and Genre Novels.' *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Santa Fe, 25 August 2018, pp. 34–43, <https://aclanthology.org/W18-4504/>.
- Van Cranenburgh, Andreas, et al. 'Vector Space Explorations of Literary Language.' *Language Resources and Evaluation*, vol. 53, 2019, pp. 625–650, doi:10.1007/s10579-018-09442-4.
- Van Cranenburgh, Andreas, and Corina Koolen. 'Results of a Single Blind Literary Taste Test with Short Anonymized Novel Fragments.' *Proceedings of The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Barcelona, 12 December 2020, pp. 121–126, <https://aclanthology.org/2020.latechclfl-1.14.pdf>.
- Van der Deijl, Lucas, et al. 'Mapping the Demographic Landscape of Characters in Recent Dutch Prose: A Quantitative Approach to Literary Representation'. *Journal of Dutch Literature*, vol. 7, no. 1, April 2016, <https://www.journalofdutchliterature.org/index.php/jdl/article/view/103>.
- Van Rossum, Lisanne M., et al. 'I Catching: Computationally Operationalising Narrative Perspective for Stylometric Analysis'. *Digital Humanities Benelux Journal*, vol. 3, 2021, pp. 1–12, doi:10.5281/zenodo.3855653.
- Van Wingerden, Wouter, and Pepijn Hendriks. '*Dat hoor je mij niet zeggen!*' [*You Will Not Catch Me Saying That*], not available in English]. Thomas Rap, 2015.

7. The Riddle of Literary Quality Solved?

Reflection and Conclusion

Abstract: The chapter sums up the main results of the project The Riddle of Literary Quality. For Dutch readers in 2013, genre fiction had no chance of literary fame. Literary novels with a more complex or difficult language, structure, or content received higher ratings. Male-authored literary novels were still more highly valued than those written by female authors and this reflects the Dutch society in which the conventions about literary quality are embedded. Being aware of underlying prejudices may lead to changes in the future and suggestions are done how to conduct follow-up research into, for example, the literary canon. More methodological research is needed into the influence of amount of dialogue, first-person or third-person perspective, and narrative tense.

Keywords: Literary quality, Literary Canon, Literary Bias, Computational Literary Studies, Distant Reading

The Computational Literary Studies project The Riddle of Literary Quality that ran in Amsterdam from 2012 to 2019 posed a challenging question: Is literary quality measurable? In search of an answer to this controversial question the project brought together two different approaches. The first was a large survey conducted in the Netherlands in 2013, The National Reader Survey. It gathered actual readers' opinions about the general quality (how 'good' the book was perceived to be) and literary quality of 401 novels published in Dutch for the first time in 2007 or later. The second was a computational analysis of the full digital texts of these novels. Correlating the results of both approaches was expected to deepen existing knowledge about which textual and extra-textual features could play a role in the perceptions of literariness. In this closing chapter I will briefly recap the main conclusions, with a reference to the chapters in which I addressed them. I will also expand on some observations and reflect on their consequences for future work.

Conventions about literary quality are embedded in the society from which they arise. The research therefore was deliberately limited to the Netherlands, a

Western European country that in 2013 had nearly seventeen million inhabitants. Its official language, Dutch, is also the language spoken in Flanders, which is part of the Netherlands' neighbour to the south Belgium (Ch. 1). A couple of hundreds of participants from Belgium shared their opinions in The National Reader Survey, but most of the 13,784 respondents came from the Netherlands and their take on literary quality can be seen as representative for the Dutch reading, book-buying, and book-borrowing audience at the time (Ch. 2). The Netherlands plays a role in the periphery of the 'world republic of letters' dominated by the English language and culture. Literary fields in the periphery usually are characterized by a large proportion of translations, and the Netherlands is no exception. The abundance of translations on the Dutch market is a fact that research into literary perceptions should take into account. In 2009, translations from other languages into Dutch in the categories of fiction and poetry amounted to 65 to 70 per cent (Ch. 1). In our corpus of 401 novels, 249 were translations, which is 62 per cent.

The aim of the project was to analyse perceptions of literariness on a large scale: not only considering the opinions of scholars, teachers, reviewers, publishers, and prize juries, but more prominently that of a large group of readers and non-readers. The opinions were gathered in the form of ratings of novels read and unread, and motivations of the literary quality rating for one of the novels a participant had read and rated. The amount of novels about which participants could share their opinion was limited to four hundred novels and one bonus title, in order to generate enough ratings per book for statistically sound observations. The books were selected based on a set of strict criteria: novels or separately published stories of at least twenty pages published for the first time in Dutch in 2007 or later; most sold in bookstores and most borrowed from public libraries in 2010–2012; and not limited to 'literary' fiction but including 'genre' fiction and young adult fiction. Furthermore, the selection was based on the genre labels attributed to the novels by publishers. Following these criteria led to a research corpus that included titles from three main categories: 147 Literary novels (37 per cent of our corpus), 186 Suspense novels (46 per cent), 41 Romance novels (10 per cent), and a miscellaneous group of 27 books (Other, 7 per cent) with different labels such as a few regional novels, young adult fiction, fantasy, and a couple of literary non-fiction labelled as novels by their publishers (Ch. 2).

This set of criteria made the selection of titles verifiable and replicable – other scholars following the same rules are expected to end up with more or less the same list of books. However, the resulting corpus definitely was not balanced in any way. It did not adequately represent all fiction published in the three years before the survey, but only reflected which novels were most popular in the bookstores and public libraries in the Netherlands during these three years. One of the concerns was that the corpus would not include enough highly rated, 'highbrow', literary

novels for a comparison with lowly rated books. Luckily, this turned out not to be a problem: the spread of mean scores for literary quality was wide enough for relevant observations (Ch. 2). It needs to be emphasized, however, that the corpus did not contain books that only a small, select group of readers considered to be highly literary. Therefore, the observations gathered for the corpus, and the conclusions drawn, may not necessarily be transferable to *all* literary novels. It remains unclear what the differences are between best-selling novels that are perceived as being highly literary, and other highly literary novels that are being read and appreciated by a much smaller readership. Follow-up research is required.

The selection of computational text analysis measures to be applied was also informed by their potential verifiability and replicability. They ranged from low-level, easy to reproduce measures such as word frequencies and keyword analysis, to high-level measures dependent on more metadata and more complex coding. Examples of such high-level measures are sentence complexity based on dependency parsing, and predictive modelling. A major drawback is that for legal reasons the digital corpus of full texts cannot be shared with others. To overcome this issue, as many data as can be legally shared are presented on Github and in the litRiddle package for R (see appendix 3 for more details). In the R package, a substantial part of the novels' word frequency lists can be found, so that most of the bag-of-words approaches can be replicated and verified. I hope that somewhere in the future of the digital scholarly infrastructures that are currently being built, scholars will be allowed access in controlled environments to full texts still in copyright, for both new research and for verification and replication of previous research. With this monograph, I intend to illustrate how important such access is for innovative and relevant research in the Humanities, and I hope it will advance such developments.

The results of high-level measurements, mostly done by Andreas van Cranenburgh, were discussed in Chapter 6 and will be summarized farther on in this concluding chapter. My own experiments were more low-level and made use of freely available and accessible software. I used AntConc for keyword analysis to contrast the vocabularies of different sets of novels or readers' motivations, and the Stylo package for R for Delta-based cluster analysis and principal components analysis to find out which novels were most similar or different in their word frequencies (Ch. 2–6). The results of these analyses did not fully explain differences in perceived literary quality, but did help to find out which next analytical steps to take. They also led to some methodologically interesting observations and ideas for follow-up research. Continually, the biggest differences between novels rated with high scores for literary quality and low ones seemed to revolve around the use of present tense or past tense, and an emphasis of first-person versus third-person pronouns. This use of pronouns could be related to a predominance of a first-person or third-person narration, but may also interfere with the amount of dialogue in a novel, which

is expected to invite more first-person pronouns (Ch. 4, 5, and recapped in Ch. 6). Andreas van Cranenburgh showed that novels with low scores for literary quality usually have more dialogue (Ch. 6), but it may well be that additionally also tense and narrative perspective do play a role in the 2013 conventions of literary quality. However, at this moment these observations will have to remain mere suggestions. For these measurements I used very small sets of novels, that only represented the most widely sold and borrowed books of which the research corpus consisted. The results, however preliminary they may be, do help to think about the applied measures themselves. Based on the many principal components analyses of Dutch novels that I made, I come to the overarching hypothesis that the characteristics of the Dutch language expressing tense and narrative perspective may for a large part explain what in the PCAs is found in the first and second component. This hypothesis definitely needs further exploration, for instance with experiments with manipulated text samples. Such follow-up research would show whether my assumption is correct for the Dutch language. Only for Dutch, because different languages express narrative tense and focalization in linguistically different ways. Other languages would need their own experiments, that may be compared with and contrasted to the experiments with Dutch novels. This could help to make the still enigmatic method of PCA more transparent and explicable for computational literary studies.

The innovations of The Riddle of Literary Quality project are manifold. Apart from the novel combination of a large-scale reader survey with computational text analysis, the inclusion of both translations and genre fiction (in this case mostly Suspense and Romance) is unique. Let me return to Karel van het Reve's famous Huizinga talk from 1978, from which I quoted in the first chapter of this book:

One of the key questions of literary theory should really be: Is it possible to make a description of a good book, without using the word 'good', which is not applicable to a worthless book? And vice versa: Is it possible to describe a bad book, again without using the word 'bad' or similar words, in such a way that it is impossible to make or find a good book that fully fits that description? This crucial and interesting question is avoided like the plague in literary theory. This question has been eliminated in one fell swoop by placing poor-quality literature outside of literature, by excluding from research one of the most interesting phenomena of literature, difference in quality. (Van het Reve 135–136, my translation; cf. Van Dalen-Oskam [2021] 18–19)

The Riddle approach led to insights into relationships and contrasts in terms of genre labelling, stylistic properties, and reader opinions about literary quality. The motivations that readers provided show that literary quality is understood in terms

of writing style and the so-called 'deeper layers' in the works. Explanations that go beyond the conceptual level were not given. Readers also often referred to a novel's genre to motivate their score for literary quality. The influence of genre in Dutch perceptions of literariness was confirmed by the clear differences in scores of the main categories in our corpus, revealing a genre hierarchy with Literary novels at the top, Romance at the bottom, and Suspense in between (Ch. 2).

Novels and authors crossing genre borders can be confusing for both publishers and readers. Authors, when writing, usually focus on a certain genre, position themselves as authors of suspenseful, romantic, or literary books and, consciously or unconsciously, largely adopt the style that they perceive as 'fitting' within their preferred genre – close enough, at least, to the style used in existing novels from that category. In this way, genre characteristics are largely maintained. And publishers, reviewers, and readers approach the text in a similar way. Changes in genres generally occur only very gradually, with a few exceptions (Ch. 6). One of these exceptions is the *Vijftig tinten* (*Fifty Shades*) trilogy. Its publication has resulted in a great deal of economic capital, but hardly any symbolic capital – literary appreciation – for the author (Ch. 3). It did substantially change the wider literary field, however.

Another exception is the work of Dutch-language author Arnon Grunberg. In the text analysis experiments, Grunberg's two highly rated literary novels closely resembled books from the Romance category, representing the genre which got the lowest scores for literary quality. In my estimation, no Grunberg reader would even consider assigning that genre label to his novels. Books labelled Romance generally presented easy-to-read stories about careers and relationships, had a light-hearted tone despite the conflicts they contained, and usually ended well. Grunberg's novels never end happy, but their unusually 'transparent' style for a literary novel draws extra attention. The relatively short sentences, simple vocabulary, and the everyday subject matter – perhaps precisely because they evoke an association with books from the Romance category – make the gruesome events and bad endings in Grunberg's novels all the more striking. The deviation from the style that readers expect from a writer known as 'literary' has thus become a strength rather than a weakness for the literary standing of Grunberg's work. If we had not included Romance novels into the equation, we would not have been able to observe this so clearly (Ch. 5).

Literary scholars have always had different views on what makes a text literary. At one end of the spectrum are researchers who believe that quality lies in the text itself, with its use of surprising and deviating language, and at the other end are researchers who mainly attribute literary quality to sociological processes dictated by the (unconscious) consensus of the cultural elite. It seems that Karel van het Reve adhered to the first view, but he did not present an answer to the question that is so central to this book. Even without it, Van het Reve himself and his 1978

lecture have made history for highly educated readers with a relevant educational background. Some participants in The National Reader Survey even referred to the Huizinga lecture. For example, when prompted to explain why he gave Lee Child's suspense novel *De affaire (The Affair)* a 2, the second-lowest score, for literary quality, a 61-year-old male reader replied:

No. Karel van het Reve already asked that question in his brilliant (literary!) Huizinga lecture (1978). And I can assure you: you cannot answer the question of what literature is and what is not. First solve the question Karel van het Reve raised back then: why is one book very readable for one reader and not at all for another? Incidentally, I find it absolutely incomprehensible that a scientist, after 1978, still dares to ponder the question of what literature is and what is not. Nothing learned from that Huizinga lecture? The memory of literary scholars is like a sieve. Lee Child's writing is 'supple' and 'suspenseful'. But so is F. Springer's and Murakami's writing. So what? (ID_2282)

Although the respondent misinterprets Van het Reve's question, his own question is equally relevant: 'why is one book very readable for one reader and not at all for another'? My answer to that question is that the difference lies in the readers. Not every reader is looking for the same reading experience. Virtually all readers in one of our representative test panels wanted to be carried away by a book. But not all of them wanted to be challenged by a book to go in search of deeper layers, which is precisely what 'literary' readers feel they are supposed to do. Literary reading seems to be an extra step in a reader's development. Literary readers begin as 'pleasure readers'. But pleasure readers need not always develop into literary readers (Ch. 2). A more personal answer to this participant would be: that you not only find the novels of thriller author Lee Child enjoyable and exciting, but also the literary novels of F. Springer and Haruki Murakami, suggests that you are a 'literary reader'. Unfortunately, there were not enough readers among the participants in The National Reader Survey who were primarily pleasure readers to investigate whether readers who read the rather literary novels by Springer and Murakami give a similar assessment of the literary quality of Child's Suspense novels as readers who only read for pleasure, or were more strict with the work of Child.

What we *did* find out is that literary conventions in the Netherlands in 2013 could be linked, on the one hand, to textual and content-related characteristics and, on the other, to the way in which professional and non-professional readers exchanged views on literature. The field theory of Bourdieu, which in fact was explicitly challenged in this research, thus could not be discarded: sociological processes in which prestige – symbolic capital in Bourdieu's terms – is distributed or withheld proved to be a serious explanation for the biases we noticed in the

dominant literary perceptions. An additional outcome was that those biases were to a large extent measurable using predictive modelling. However, the mixture of text and context was different than most researchers thought so far. The influence of the text itself was greater than expected by those focused on Bourdieu and his sociology of art, yet also smaller than those who advocate for the purely intrinsic aesthetics of literary language (Ch. 6). In fact, the answer lies somewhere in-between both extremes, and as such aligns with the ideas expressed by researchers of the new sociology of art, that aimed to bring back the objects of art into sociological research (Ch. 1).

The text analysis experiments have shown that the language in highly literary novels usually was more complex. Textual complexity has many variables, of which we have only investigated the most measurable and replicable ones for now. Especially sentence length (the mean number of words per sentence per novel) turned out to be a rather good predictor, although we could establish this mainly for Literary novels written by male authors. The Literary novels by Dutch-language female authors in our corpus did not show equally long sentences. As stated in this book (Ch. 6), this does not prove that female authors never used such long sentences, but can only lead to the observation that such novels did not make it to the bestseller list or the top of the list of books most borrowed from public libraries that we used to select the novels for our corpus of 401 books. In follow-up research we therefore need to broaden the research by adding less well-known novels from the same time period to the corpus, to explore which ratings would be attributed to novels by female authors with a mean sentence length that compares to those of the highly valued novels by male authors in our corpus.

Furthermore, highly valued novels showed more formal language and had a higher number of topics. Content-wise it seemed that associations with femininity reduced the likelihood of literary appreciation by readers, *in particular* for female authors (Ch. 6). In the Dutch literary field of 2013, therefore, it seems that masculinity was rewarded. Except when authors or their work could be associated with a popular sport, as we saw for author and former sports journalist Mart Smeets.

The differences in appreciation of literary novels written by female and male authors were obvious. This fact provided a unified picture of the literary conventions prevailing in the Netherlands in 2013. The participants in The National Reader Survey were representative of the public that read (literary) fiction and that audience clearly took male authors more seriously than female authors when it came to literary appreciation (Ch. 2, Ch. 6). This – probably mostly unconscious – bias was not unique to the literary field. It is a reflection of what we also saw in many other parts of Dutch society and makes clear that opinions about literature are not isolated but are embedded in and generated by society as a whole. Judgements about literary value can therefore also be taken as an index of prejudices in society at

large that otherwise may be ignored or stay unnoticed. In sum, the riddle of literary quality has not been solved, but its nature has changed: how to explain that such preconceptions, possibly unconsciously, play a role in the literary appreciation of novels? The provisional answer is that readers, publishers, and authors have internalized these associations by looking at what others have done and thought before them. Consciously or unconsciously, they copy existing conventions and propel them forward. Such conventions may already have a long history, in which changes have gradually been brought about. There is still a lot of research to be done into how exactly this process of continuation and change works, and for this we need to go back in time, among other things. How objective, for example, is the literary canon if we historically contextualize it? We need to include forgotten novels in such research for a better window into the past to learn how unconscious and conscious prejudices influenced which books we still know and appreciate today. I hope to tackle the riddle of the Dutch literary canon in a subsequent book.

All in all, attributing literary value to novels is a process full of mutual dependencies and may be seen as a complicated feedback loop. From all the manuscripts offered to publishers, editors select for publication those novels expected to bring in profit when they are labelled as literary novel (Ch. 6). The profit can be economic (money) or symbolic (prestige). This selection process takes place in reference to existing publications, so that editors and publishers largely confirm and perpetuate prevailing conventions. Publishers must strike a balance between innovation and deviation. The more a manuscript deviates from what is already available, the greater the risk to the publisher. Still, publishers take risks regularly, hoping that a breakthrough title is profitable enough to keep publishing those other books.

Authors, too, play a role in this system of mutual dependencies. They are conscious of the current literary landscape and when they are writing, they have in mind which part of that market they would like to break into, and how they want to position themselves as authors. It informs the choices they make (Ch. 6). After all, they want to be read, otherwise they would not submit their manuscripts to publishers in the hope of getting them published as a 'real book'. Authors instinctively know which subjects are appropriate for a literary novel and which they would do better to avoid, and which writing style is most appropriate to that end. And it seems that in the Netherlands in 2013, male authors had more freedom of choice of style and subject matter without losing prestige than female authors. In this way, published authors largely reinforce existing conventions. Authors, too, take risks, and as with publishers, some risks pay off in great literary prestige (and enough buying or borrowing readers and perhaps prize money to guarantee a liveable wage) while others do not have the desired effects of relevance and acclaim.

Every writer starts out as a reader. Readers too are part of the system of mutual dependencies. As children, they are given books to read which adults have selected

for them. From that starting point, many children choose their own reading path, in which an essential role is played by public libraries filled with books selected for children – by adults. In Dutch secondary school, pupils are presented with a reading list compiled by teachers. That reading list often reflects the choices made by older generations of teachers and upholds existing conventions about what constitutes literature. The main message Dutch pupils seem to get is: literature must be difficult. And the second message is that apparently men are better at writing literature than women, because there are many, many more male authors on most reading lists than female authors.¹ Anecdotal evidence shows many Dutch children start to hate obligatory reading during their secondary school years and stop as soon as they can. Only the persistent ones read on, and doing so, they keep the existing attitudes alive. And as adults, they are guided in their reading choices by the reading choices of other avid readers: friends, book reviewers, booksellers in shops and on television, and literature researchers, who, like them, largely uphold existing ideas about what literary quality entails.

And so we come full circle. The shortest summary is that in the 2013 Dutch literary field, the literary quality of a novel was most often linked to the degree of difficulty of a book. To take this a step further: good literary writers were intelligent writers and intelligent writers were more likely to be male than female. I am convinced that publishers, authors, and readers are generally unaware of this process and of their role in it. And that few literary readers realize that, by labelling a book as having high literary quality they are not rating the book, but they are judging the reader. The reader who is intelligent enough to read, understand and appreciate complex texts, and who can make conversation about them or perhaps even write about them. The reader who has difficulty reading, but has the opportunity to practise a lot and thus gain a lot of reading experience. The reader who can spend money on buying books or on a subscription to the public library. The reader who, after their daily routine, still has energy left to read. Or who has a quiet place in their home to isolate themselves from an environment that is not reader-friendly.

How does the Dutch situation compare to that in other countries? Every year, the Organisation for Economic Co-operation and Development (OECD) in their Programme for International Student Assessment (PISA) measures, as their website states, '15-year-olds' ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges'. Since the year 2000, the reading ability of Dutch children is steadily trending down, although they do still score on the highest possible level. Many Dutch scholars and teachers are concerned about this

¹ The imbalance in the reading lists has been convincingly shown by Elisa Janssen in her master's thesis. A slightly different perspective on the gender imbalance is discussed by Dera 2018, 157 and Dera 2019, 4 and *passim*.

'reading crisis', as it has been dubbed, and are developing plans how to address the issue. Opinions on how to best approach this differ.² The same goes for the reasons for this decline. Based on the results of the project *The Riddle of Literary Quality*, I would like to propose the hypothesis that the difficulty that plays such an important role in the Dutch literary conventions is what intimidates beginning readers and in fact pushes more and more students away from reading at all. To test this hypothesis, we would need to find out what the conventions are in other countries, for example in Estonia and Portugal, which both show a steady increase in reading ability over the same years and in 2018 outperformed the Netherlands.³ Is the concept of 'highbrow' literature more important in the Netherlands than in these two countries? And how does this compare to the United States, Canada, or the United Kingdom, which rank higher than the Netherlands? I have the impression that in these countries, the prestige of highbrow literature is emphasized far less than in the Netherlands. Is that indeed the case?

A first step towards an international comparison is the project *Novel Perceptions*. At the University of Wolverhampton, Professor Bas Groes and I work together with a whole team of researchers to replicate the Riddle research in the United Kingdom, with funding from the Arts and Humanities Research Council. At the time of writing this conclusion, the *Big Book Review* – moulded in the image of the Dutch 2013 National Reader Survey – is still in full swing, with a list of four hundred novels for which readers' opinions are being gathered. The first results of the 'UK Riddle' are expected to be published in 2025. And I hope other countries will follow. A comparison of literary conventions in different countries not only is relevant from the perspective of reading ability. It is equally important to find out if the gender bias in the Dutch literary field also shows up elsewhere, or whether other kinds of biases can be uncovered. This kind of research may contribute to awareness about such issues. Although it is extremely difficult to break away from existing conventions, being aware of underlying prejudices may lead to changes in the future.

In the years since 2013 when *The National Reader Survey* was held in the Netherlands, many new female literary authors have published work to great acclaim. In response to the results of *The Riddle of Literary Quality*, people have often remarked that the contemporary situation seems to be very different for female authors. However, I am not yet convinced. As Corina Koolen mentioned in her PhD-thesis, this kind of remarks can be found in past decades as well – and upon further

2 As can be seen in a 2023 volume of contributions addressing different perspectives on this 'reading crisis' edited by Yra van Dijk et al.

3 Background information about these countries' improved PISA scores can be found in a 2022 report of the European Commission, *Successful PISA Stories in the EU: How Some Member States Have Been Able to Improve Their Performance over Time*.

scrutiny, nothing really did change. The Riddle project did develop a way to test such views. And by repeating the reader survey in later years, we may actually be able to monitor the effects of the resulting discussions and view some actual changes in value attribution. I wish for a repeat National Reader Survey in the Netherlands to be organized in a couple of years so we may learn more about the way in which today's perceptions of literary quality become embedded in memories and minds of next generations of readers and materialize in the reading lists of the future.

References

All hyperlinks were last accessed in February 2023.

- Bourdieu, Pierre. *The Field of Cultural Production: Essays on Art and Literature*, edited by Randal Johnson. Polity, 1993.
- Dera, Jeroen. 'De lezende leraar: Literatuuronderwijs in Nederland(s) als onderzoeksobject' ['The Reading Teacher: Dutch Literary Education in the Netherlands as an Object of Research', not available in English]. *Tijdschrift voor Nederlandse Taal- en Letterkunde*, vol. 134, no. 2, 2018, pp. 146–170.
- Dera, Jeroen. *De praktijk van de leeslijst: Een onderzoek naar de inhoud en waardering van literatuurlijsten voor het schoolvak Nederlands op havo en vwo* [*The Practice of the Reading List: A Study of the Content and Appreciation of Reading Lists for Dutch as School Subject in Senior General Secondary Education and Pre-University Education*, not available in English]. Stichting Lezen, 2019, https://www.lezen.nl/sites/default/files/praktijk_van_de_leeslijst_digitaal.pdf.
- European Commission, Directorate-General for Education, Youth, Sport and Culture, *Successful PISA stories in the EU: how some Member states have been able to improve their performance over time*, final report, Publications Office of the European Union, 2022, <https://data.europa.eu/doi/10.2766/429517>.
- Janssen, Elisa. 'We hebben wel heel veel mannen, laten we nu eens een vrouw nemen': Een verkennend onderzoek naar een genderkloof in het literatuuronderwijs in de bovenbouw van havo en vwo. ['We Have Rather a Lot of Men, Let's Add a Woman for a Change: An Exploratory Study into a Gender Gap in Literature Education in Advanced Stage Senior General Secondary Education and Pre-University Education,' not available in English]. 2019. University of Amsterdam, master's thesis. *University of Amsterdam Digital Academic Repository*, https://scripties.uba.uva.nl/search?id=record_23761.
- Koolen, Corina. *Reading Beyond the Female: The Relationship Between Perception of Author Gender and Literary Quality*. 2018. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37f1ce7>.

- 'Novel Perceptions: Towards an Inclusive Canon Project Website.' *University of Wolverhampton*, <http://novel-perceptions.thememorynetwork.com/>.
- 'Organisation for Economic Co-operation and Development (OECD).' *Programme for International Student Assessment (PISA)*, <https://www.oecd.org/pisa/>.
- Van Cranenburgh, Andreas. *Rich Statistical Parsing and Literary Language*. 2016. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37fice7>.
- Van Dalen-Oskam, Karina. *Het raadsel literatuur: Is literaire kwaliteit meetbaar? [The Riddle of Literature: Can We Measure Literary Quality?]* Amsterdam University Press, 2021, <https://karinavdo.github.io/RaadselLiteratuur/>.
- Van Dijk, Yra, et al. *Omdat lezen loont. Op naar effectief leesonderwijs in Nederland [Because Reading is Beneficial. Towards Effective Reading Pedagogy in the Netherlands]*, not available in English]. Pica, 2022.
- Van het Reve, Karel. 'Literatuurwetenschap: het raadsel der onleesbaarheid' ['Literary Theory: The Riddle of Illegibility,' not available in English]. Johan Huizinga Lecture, 8 December 1978, Leiden, https://www.dbnl.org/tekst/reve003lite02_01/index.php.

Appendix 1: The Survey

1. What is your age?

Note: Value between 1 and 110; under 16: to end.

2. How many books do you read on average per year? If you do not know the exact number, you can make an estimate.

Note: If a participant entered 0, they were not shown questions 4 to 8 and question 12.

3. What sort of books do you read?

1. Exclusively fiction (novels, thrillers, fantasy, etc.)
2. Exclusively non-fiction (business, hobby, biographies, study books, etc.)
3. Both fiction and non-fiction

Note: If a participant selected option 2, they were not shown questions 4 to 8 and question 12.

4. You will find a number of statements below. Could you indicate to which extent you agree or disagree with these statements?

Scale: strongly disagree (1), disagree (2), do not agree/do not disagree (3), agree (4), strongly agree (5), do not know (6)

The items were presented in random order.

- 1 I like to read novels that I can relate to my own life
- 2 I am mainly concerned with the story in the novel
- 3 The writing style in a book is of importance to me
- 4 I like to find the deeper layers of a novel
- 5 I like reading literature
- 6 I read novels to discover new worlds and unknown ages
- 7 I mainly read novels during holidays
- 8 I read several novels at the same time

5. Below you will find a list of 401 books. We would like to ask you to select the books you have read. You can browse the list, but you can also search for title or author. You can select as many as you like. You can check as many books as you have read, but you can also make a smaller selection. After this, you will receive a random list of seven books you have checked for a more specific judgement. You can then choose if you want to judge more novels.

Note: The list of 401 books can be found in Appendix 2.

- 6. Below you will find books you have read. Could you judge these books? To which extent did you think the book was good or bad? We would like to know your personal opinion.**

Note: A maximum of seven read books were shown, randomly selected from those ticked as read. Scale:

1. *Very bad*
2. *Bad*
3. *Somewhat bad*
4. *Not good, not bad*
5. *Somewhat good*
6. *Good*
7. *Very good*
8. *Do not know¹*

- 7. Below, you will again find the books that you have read. Could you judge these books for their literary value? To which extent did you think this book was literature? We would like to know your personal opinion.**

Note: The same books were shown here, in the same order, as under question 6.

Scale:

1. *Definitely not literary*
2. *Not literary*
3. *Tending towards non-literary*
4. *Bordering literary and non-literary*
5. *Somewhat literary*
6. *Literary*
7. *Highly literary*
8. *Do not know*

You can now pick:

1. **If you want to continue the list of questions (about 3 minutes in total)**
2. **If you want to judge seven more books that you have read (about 6 minutes in total; you will make us extra happy)**

- 8. Could you explain why you chose this judgement?**

Note: The participant was presented with the author and title of the book for which the person had given the highest or lowest score in question 7. In an open field, the participant could substantiate the score in their own words.

¹ In the survey, for questions 6 and 7 and 10 and 11, we used the reversed values, i.e. (1) very good–(7) very bad, and (1) highly literary–(7) definitely not literary. For our analysis, we have re-scored the results. In this book, I consistently use the re-scored values.

9. Below, you will find a list of four hundred books. Are there books here that you have not read, but do have an opinion about? For instance, because you expect the book to be good or bad. Could you select the books from this list that you have not read, but that you do have ideas about?

Note: In this list, the books previously marked as read by the participant were not shown.

10. Below you will find the books that you have not read but do have an opinion about. Could you judge these books? To which extent do you expect you would find this book good or bad?

Note: A maximum of seven unread books were shown, randomly selected from those ticked in question 9.

Scale:

1. *Very bad*
2. *Bad*
3. *Somewhat bad*
4. *Not good, not bad*
5. *Somewhat good*
6. *Good*
7. *Very good*
8. *Do not know*

11. To which extent do you expect these books to have literary value?

Note: The same books were shown here, in the same order, as under question 10.

Scale:

1. *Definitely not literary*
2. *Not literary*
3. *Tending towards non-literary*
4. *Bordering literary and non-literary*
5. *Somewhat literary*
6. *Literary*
7. *Highly literary*
8. *Do not know*

- 12a. On a scale of 1–7, where 1 means a non-literary reader and 7 means a very literary reader, where would you place yourself as a reader?

Non-literary reader 1—2—3—4—5—6—7 very literary reader

12b. Below you will find a number of statements. Could you indicate to which extent you agree or disagree with these statements?

Scale: strongly disagree (1), disagree (2), do not agree/do not disagree (3), agree (4), strongly agree (5), do not know (6)

The items were presented in random order.

- 9 I like novels based on true stories
- 10 I like to think about the plot structure of a novel
- 11 The writing style in a novel is more important to me than the story
- 12 I want to be swept away by a novel
- 13 I like to choose my books from the top 10 lists of bestsellers
- 14 I read novels in order to be challenged intellectually
- 15 I like novels that are easy to read
- 16 In the evening, I prefer reading a book over watching television

13. Are you

1. Male
2. Female
3. No answer

14. What is your highest obtained level of education or which level are you pursuing at the moment?

1. No education / primary school
2. Vocational training
3. Vocational training (theoretical/mixed)
4. Secondary school
5. Community college
6. College
7. Graduate school
8. No answer

15. What are the four numbers of your postal code?

16. If you have any remarks on the survey or the answers you gave, you can make them here.

Note: In an open field, the participant could add any other remarks about the survey.

Appendix 2: The Books

The 401 novels from The National Reader Survey are sorted here according to their mean score for literary quality, from high to low. The column on the far right shows the mean score for general quality.

*Books that received fewer than fifty reader ratings.

†Books that contain one longish story or a collection of shorter stories or essays and have therefore not been included in some analyses.

Rank	Author	Title ¹	Literary quality	General quality
1	Julian Barnes	<i>Alsof het voorbij is</i>	6.62	6.54
2	Erwin Mortier	<i>Godenslaap</i>	6.60	6.16
3	Erwin Mortier	<i>Gestameld liedboek</i>	6.54	6.48
4	Michel Houellebecq	<i>De kaart en het gebied</i>	6.41	6.14
5	Tom Lanoye	<i>Sprakeloos</i>	6.37	6.43
6	Haruki Murakami	<i>Norwegian Wood</i>	6.31	6.28
7	A.F.Th. van der Heijden	<i>Tonio</i>	6.29	6.31
8	Stephan Enter	<i>Grip</i>	6.26	6.11
9	J. Bernlef	<i>Geleende levens</i>	6.25	6.23
10	Umberto Eco	<i>De begraafplaats van Praag</i>	6.24	5.84
11	Tommy Wieringa	<i>Caesarion</i>	6.20	5.90
12	David Mitchell	<i>De niet verhoorde gebeden van Jacob de Zoet</i>	6.19	6.38
13	Gerbrand Bakker	<i>De omweg</i>	6.19	6.18
14	Haruki Murakami	<i>1q84</i>	6.19	6.22
15	Arthur Japin	<i>De overgave</i>	6.17	6.25
16	Thomas Rosenboom	<i>Zoete mond</i>	6.17	5.78
17	Thomas Rosenboom	<i>Mechanica</i> †	6.16	5.88
18	Arnon Grunberg	<i>Huid en haar</i>	6.16	5.90
19	Jonathan Franzen	<i>De correcties</i>	6.16	6.09
20	Jonathan Franzen	<i>Vrijheid</i>	6.15	6.08
21	J. Bernlef	<i>De een zijn dood</i>	6.14	6.05
22	Arnon Grunberg	<i>Selmonosky's droom</i> †	6.13	5.86
23	Arthur Japin	<i>Vaslav</i>	6.11	6.04
24	Arnon Grunberg	<i>De man zonder ziekte</i>	6.11	5.84
25	Stefan Brijs	<i>Post voor mevrouw Bromley</i>	6.09	6.21

¹ Whenever a title is mentioned in this book, the English title (when available) or a provisional translation into English is added (cf. the Index to locate these). The English titles for all 401 books can be found through the companion website in csv files and in the litRiddle package for R (see Appendix 3).

Rank	Author	Title	Literary quality	General quality
26	P.F. Thomése	<i>De weldoener</i>	6.09	5.69
27	Adriaan van Dis	<i>Tikkop</i>	6.07	5.97
28	J.J. Voskuil	<i>De buurman</i>	6.05	6.17
29	Tommy Wieringa	<i>Portret van een heer †</i>	6.04	6.03
30	Arthur Japin	<i>Dooi †</i>	6.03	6.01
31	Laurent Binet	<i>HhhH</i>	6.02	6.29
32	Nicole Krauss	<i>Het grote huis</i>	5.99	5.97
33	Bernard Dewulf	<i>Kleine dagen</i>	5.98	5.72
34	F. Springer	<i>Quadriga</i>	5.95	5.98
35	Jan Siebelink	<i>Het lichaam van Clara</i>	5.94	5.71
36	Margriet de Moor	<i>De schilder en het meisje</i>	5.90	5.76
37	John Irving	<i>In een mens</i>	5.88	6.04
38	Chad Harbach	<i>De kunst van het veldspel</i>	5.88	6.20
39	Paolo Giordano	<i>De eenzaamheid van de priemgetallen</i>	5.88	5.90
40	Maarten 't Hart	<i>Verlovingstijd</i>	5.86	5.92
41	Charles Lewinsky	<i>De verborgen geschiedenis van Courtillon</i>	5.86	5.93
42	Marente de Moor	<i>De Nederlandse maagd</i>	5.86	5.23
43	Peter Buwalda	<i>Bonita Avenue</i>	5.84	5.82
44	Kader Abdolah	<i>De koning</i>	5.84	5.91
45	Tom Lanoye	<i>Heldere hemel</i>	5.83	5.71
46	Dimitri Verhulst	<i>De laatste liefde van mijn moeder</i>	5.82	5.93
47	Anna Enquist	<i>De verdoovers</i>	5.81	5.80
48	Jan Siebelink	<i>Oscar</i>	5.81	5.55
49	Carlos Ruiz Zafón	<i>De gevangene van de hemel</i>	5.80	6.21
50	John Irving	<i>De laatste nacht in Twisted River</i>	5.77	5.97
51	Kees van Beijnum	<i>Een soort familie</i>	5.76	5.94
52	Leon de Winter	<i>Recht op terugkeer</i>	5.75	5.88
53	Julia Franck	<i>De middagvrouw</i>	5.73	5.90
54	Charles Lewinsky	<i>Het lot van de familie Meijer</i>	5.72	6.14
55	Tessa de Loo	<i>De grote moeder †</i>	5.67	5.79
56	Jessica Durlacher	<i>De held</i>	5.66	5.78
57	Carlos Ruiz Zafón	<i>Het spel van de engel</i>	5.66	5.98
58	Isabel Allende	<i>Het eiland onder de zee</i>	5.64	6.09
59	Dave Eggers	<i>Wat is de wat</i>	5.64	6.14
60	Irvin D. Yalom	<i>Het raadsel Spinoza</i>	5.63	6.21
61	Niccolò Ammaniti	<i>Ik en jij</i>	5.61	5.93
62	Jan Brokken	<i>Baltische zielen</i>	5.58	6.54
63	Niccolò Ammaniti	<i>Zo God het wil</i>	5.56	5.98
64	Renate Dorrestein	<i>De stiefmoeder</i>	5.53	5.81
65	Anouschka Voskuil	<i>Dorp</i>	5.52*	5.62*
66	Kader Abdolah	<i>De kraai</i>	5.51	5.23
67	Joost Zwagerman	<i>Duel</i>	5.50	5.29
68	Isabel Allende	<i>Het negende schrift van Maya</i>	5.50	6.10

Rank	Author	Title	Literary quality	General quality
69	Renate Dorrestein	<i>De zondagmiddagauto †</i>	5.49	5.69
70	Niccolò Ammaniti	<i>Het laatste oudejaar van de mensheid</i>	5.49	5.90
71	Paulo Coelho	<i>De beschermengel</i>	5.48	5.81
72	Leon de Winter	<i>VSV of daden van onbaatzuchtigheid</i>	5.47	5.44
73	Carlos Ruiz Zafón	<i>De schaduw van de wind</i>	5.46	5.82
74	Emma Donoghue	<i>Kamer</i>	5.45	6.21
75	Silvia Avallone	<i>Staal</i>	5.45	5.90
76	Karl Marlantes	<i>Matterhorn</i>	5.42	6.11
77	Khaled Hosseini	<i>Duizend schitterende zonnen</i>	5.41	6.13
78	María Dueñas	<i>Het geluid van de nacht</i>	5.37	6.14
79	Chris Cleave	<i>Kleine Bij</i>	5.36	6.10
80	Franca Treur	<i>Dorsvloer vol confetti</i>	5.36	5.65
81	Vonne van der Meer	<i>De vrouw met de sleutel</i>	5.35	5.76
82	Ronald Giphart	<i>IJsland</i>	5.34	5.36
83	Remco Campert	<i>Dagboek van een poes</i>	5.33	5.87
84	Niccolò Ammaniti	<i>Laat het feest beginnen</i>	5.33	5.64
85	Lawrence Hill	<i>Het negerboek</i>	5.32	6.33
86	Paulo Coelho	<i>Aleph</i>	5.30	5.53
87	John Boyne	<i>Het winterpaleis</i>	5.27	6.07
88	Ildefonso Falcones	<i>De hand van Fatima</i>	5.17	5.88
89	Christopher John Sansom	<i>Winter in Madrid</i>	5.13	5.92
90	Herman Koch	<i>Het diner</i>	5.12	5.45
91	Joshua Zwaan	<i>Parnassia</i>	5.11	6.06
92	Alex van Galen	<i>Süskind</i>	5.10	5.83
93	Suzanna Jansen	<i>De pronkspiegelt</i>	5.09*	5.71*
94	Herman Koch	<i>Zomerhuis met zwembad</i>	5.09	5.46
95	Geert Mak	<i>Reizen zonder John</i>	5.06	6.11
96	John le Carré	<i>Ons soort verrader</i>	5.04	6.08
97	Vonne van der Meer	<i>Het zingen, het water, de peen †</i>	5.02	5.54
98	Henning Mankell	<i>De Daisy Sisters</i>	5.00	5.78
99	Renate Dorrestein	<i>De leesclub</i>	4.98	5.18
100	Alison Pick	<i>Donderdagkind</i>	4.97	5.87
101	Jaap Scholten	<i>Kameraad Baron</i>	4.97	5.97
102	Kees van Kooten	<i>De verrekijker</i>	4.96	5.29
103	Susan Smit	<i>Vloed</i>	4.96	5.64
104	Kathryn Stockett	<i>Een keukenmeidenroman</i>	4.95	6.12
105	Henning Mankell	<i>De geschiedenis van een gevallen engel</i>	4.92	5.97
106	Jenna Blum	<i>Het familieportret</i>	4.90	5.79
107	Hannah Shah	<i>De dochter van de imam</i>	4.89	5.95
108	Tatiana de Rosnay	<i>Het huis waar jij van hield</i>	4.89	5.52
109	Anita Amirrezvani	<i>Dochter van Isfahan</i>	4.89	5.95
110	Simone van der Vlugt	<i>Jacoba, Dochter van Holland</i>	4.86	5.77
111	Simone van der Vlugt	<i>Rode sneeuw in december</i>	4.84	6.00

Rank	Author	Title	Literary quality	General quality
112	Henning Mankell	<i>Kennedy's brein</i>	4.83	5.89
113	Tatiana de Rosnay	<i>Het appartement</i>	4.82	5.56
114	Jonas Jonasson	<i>De 100-jarige man die uit het raam klom en verdween</i>	4.81	5.69
115	Tatiana de Rosnay	<i>Kwetsbaar</i>	4.81	5.65
116	George Mastras	<i>Tranen over Kashmir</i>	4.79	5.91
117	Nico Dijkshoorn	<i>Nooit ziek geweest</i>	4.79	5.69
118	Sarah McCoy	<i>De bakkersdochter</i>	4.77	5.98
119	Henning Mankell	<i>De gekwelde man</i>	4.77	6.00
120	Amanda Hodgkinson	<i>Britannia Road 22</i>	4.77	5.73
121	Henning Mankell	<i>De Chinees</i>	4.75	5.97
122	Ravelli	<i>De Vliegenvanger</i>	4.75	5.55
123	Ernest van der Kwast	<i>Mama Tandoori</i>	4.73	5.29
124	Jenna Blum	<i>In tweestrijd</i>	4.71	5.66
125	Margaux Fragoso	<i>Tijger, tijger</i>	4.71	5.41
126	Tatiana de Rosnay	<i>Haar naam was Sarah</i>	4.67	5.69
127	Ronald Giphart	<i>Zeven jaar goede seks †</i>	4.65	4.99
128	Håkan Nesser	<i>De man zonder hond</i>	4.63	5.95
129	Johan Theorin	<i>Schemeruur</i>	4.61	6.10
130	Tatiana de Rosnay	<i>Die laatste zomer</i>	4.59	5.40
131	Sarah Blake	<i>De laatste brief</i>	4.59	5.58
132	Ken Follett	<i>Val der titanen</i>	4.57	6.21
133	Heather MacDowell / Rose MacDowell	<i>Diner voor 2</i>	4.56*	5.24*
134	Robert Vuijsje	<i>Alleen maar foute mensen †</i>	4.55	5.08
135	Arnaldur Indriðason	<i>Onderstroom</i>	4.54	6.10
136	Nicci Gerrard	<i>Het weerzien</i>	4.53	5.78
137	Arnaldur Indriðason	<i>Doodskap</i>	4.51	6.04
138	Robert Vuijsje	<i>Alleen maar nette mensen</i>	4.46	4.95
139	Stephen King	<i>22/11/1963</i>	4.43	6.08
140	Heather Gudenkauf	<i>In stilte gehuld</i>	4.41	5.87
141	David Sedaris	<i>Van je familie moet je het hebben †</i>	4.39	5.34
142	Jo Nesbø	<i>De verlosser</i>	4.38	6.08
143	Vidar Sundstøl	<i>Land van dromen</i>	4.37*	5.09*
144	René Appel	<i>Weerzin</i>	4.33	5.52
145	Danielle Trussoni	<i>Het uur van de engelen</i>	4.31*	5.41
146	Rik Launspach	<i>1953</i>	4.29	5.25
147	Charles den Tex	<i>Wachtwoord</i>	4.27	5.91
148	Deon Meyer	<i>13 uur</i>	4.27	6.13
149	Jo Nesbø	<i>Het pantserhart</i>	4.27	6.26
150	Lars Kepler	<i>Getuige</i>	4.27	6.09
151	S.J. Watson	<i>Voor ik ga slapen</i>	4.26	5.85
152	Camilla Läckberg	<i>Oorlogskind</i>	4.25	5.83

Rank	Author	Title	Literary quality	General quality
153	Nicci French	<i>Medeplichtig</i>	4.23	5.76
154	Janelle Brown	<i>Alles wat wij wilden was alles</i>	4.23*	5.10*
155	Marion Pauw	<i>Zondaarskind</i>	4.23	5.64
156	Lieneke Dijkzeul	<i>Gouden bergen †</i>	4.23	5.68
157	Nicci French	<i>Dinsdag is voorbij</i>	4.22	5.75
158	Lieneke Dijkzeul	<i>Verloren zoon</i>	4.22	5.80
159	Sophie Hannah	<i>Moederziel</i>	4.22	5.66
160	Ruth Rendell	<i>De dief – Kattenkruid!</i>	4.21	5.71
161	Jussi Adler-Olsen	<i>Het Washingtondecreet</i>	4.21	5.76
162	Dennis Lehane	<i>Gone Baby Gone</i>	4.21	6.07
163	Will Lavender	<i>Het verborgen raadsel</i>	4.21	5.07
164	Unni Lindell	<i>Honingval</i>	4.21	5.89
165	Jussi Adler-Olsen	<i>De bedrijfsterrorist</i>	4.21	5.65
166	Esther Verhoef	<i>Tegenlicht</i>	4.20	5.27
167	Jo Nesbø	<i>De sneeuwman</i>	4.19	5.94
168	Camilla Läckberg	<i>Engeleneiland</i>	4.18	5.96
169	Jussi Adler-Olsen	<i>Dossier 64</i>	4.18	6.21
170	Justin Cronin	<i>De oversteek</i>	4.18	5.86
171	Chevy Stevens	<i>Vermist</i>	4.18	5.93
172	Michael Koryta	<i>Begraven</i>	4.17	6.00
173	Jo Nesbø	<i>De schim</i>	4.17	5.95
174	Jodi Picoult	<i>Negentien minuten</i>	4.17	5.93
175	Nicholas Evans	<i>De vergeving</i>	4.16	5.57
176	Carry Slee	<i>Fatale liefde</i>	4.14	5.33
177	Karin Slaughter	<i>Versplinterd</i>	4.14	6.07
178	Nicci French	<i>Tot het voorbij is</i>	4.14	5.62
179	Lars Kepler	<i>Contract</i>	4.14	6.12
180	Charles den Tex	<i>Spijt</i>	4.13	5.45
181	Hjorth Rosenfeldt	<i>Wat verborgen is</i>	4.13	6.06
182	Stephen King	<i>Gevangen</i>	4.13	6.06
183	Loes den Hollander	<i>Krachtmeting</i>	4.12	5.82
184	René Appel	<i>Van twee kanten</i>	4.12	5.63
185	Michael Robotham	<i>Gebroken</i>	4.11	6.02
186	Ruth Newman	<i>Schaduwkant</i>	4.10	5.53
187	Karin Slaughter	<i>Genesis</i>	4.09	5.99
188	Lieneke Dijkzeul	<i>Koude lente</i>	4.09	5.64
189	Loes den Hollander	<i>Troostkind</i>	4.09	5.46
190	David Baldacci	<i>Die zomer</i>	4.09	5.59
191	Camilla Läckberg	<i>Sneeuwstorm en amandelgeur</i>	4.08	5.64
192	Camilla Läckberg	<i>Vuurtorenwachter</i>	4.08	5.81
193	Camilla Läckberg	<i>Steenhouwer</i>	4.07	5.93
194	Nicci French	<i>Blauwe maandag</i>	4.07	5.49
195	Nicci French	<i>Wat te doen als iemand sterft</i>	4.06	5.56

Rank	Author	Title	Literary quality	General quality
196	Rita Spijker	<i>Tussen zussen</i>	4.06	5.05
197	Elizabeth George	<i>Lichaam van de dood</i>	4.06	5.84
198	Stieg Larsson	<i>De vrouw die met vuur speelde</i>	4.04	6.07
199	Suzanne Vermeer	<i>Noorderlicht</i>	4.04	4.99
200	Harlan Coben	<i>Blijf dichtbij</i>	4.04	6.00
201	Jussi Adler-Olsen	<i>De noodkreet in de fles</i>	4.03	6.11
202	Jens Lapidus	<i>Bloedlink</i>	4.03	5.67
203	Simone van der Vlugt	<i>In mijn dromen</i>	4.03	5.36
204	Karin Slaughter	<i>Ongezien</i>	4.03	6.08
205	Suzanne Vermeer	<i>Zomertijd</i>	4.03	5.20
206	Tess Gerritsen	<i>Het stille meisje</i>	4.02	5.90
207	Simone van der Vlugt	<i>Het laatste offer</i>	4.02	5.37
208	Camilla Läckberg	<i>Zeemeermin</i>	4.01	5.85
209	Stephen King	<i>Eenmalige zonde †</i>	4.01	5.72
210	Esther Verhoef	<i>Erken mij</i>	4.01	5.26
211	Simone van der Vlugt	<i>Herfstlied</i>	4.00	5.45
212	Stieg Larsson	<i>Gerechtigheid</i>	4.00	6.08
213	Greetje van den Berg	<i>Ergens achteraan</i>	4.00*	5.92*
214	Ruth Newman	<i>Vleugels</i>	4.00	5.73
215	Jens Lapidus	<i>Val dood</i>	4.00	5.79
216	Camilla Läckberg	<i>Predikant</i>	3.99	5.74
217	Camilla Läckberg	<i>Zusje</i>	3.99	5.52
218	Elizabeth George	<i>Een duister vermoeden</i>	3.99	5.81
219	Karin Slaughter	<i>Onaantastbaar</i>	3.99	5.90
220	Stephen King	<i>Aardedonker, zonder sterren †</i>	3.98	5.98
221	Lisa Jackson	<i>De zevende doodzonde</i>	3.98	5.94
222	Lars Kepler	<i>Hypnose</i>	3.98	5.61
223	Esther Verhoef	<i>Alles te verliezen</i>	3.97	5.42
224	Marion Pauw	<i>Daglicht</i>	3.97	5.39
225	Loes den Hollander	<i>De kat op zolder †</i>	3.97	5.63
226	Jussi Adler-Olsen	<i>De fazantenmoordenaars</i>	3.96	6.02
227	Simone van der Vlugt	<i>Blauw water</i>	3.94	5.31
228	Yvonne Kroonenberg	<i>De familieblues †</i>	3.93	5.30
229	Roel Janssen	<i>De tiende vrouw</i>	3.93	5.22
230	Tess Gerritsen	<i>Het aandenken</i>	3.93	6.03
231	Loes den Hollander	<i>Het scherventapijt</i>	3.92	5.64
232	Tami Hoag	<i>Dieper dan de doden</i>	3.91	5.85
233	John Grisham	<i>De wettelozen</i>	3.91	5.78
234	Loes den Hollander	<i>Zielsverwanten</i>	3.91	5.51
235	Loes den Hollander	<i>Vluchtgedrag</i>	3.90	5.78
236	Harlan Coben	<i>Levenslijn</i>	3.90	5.92
237	Stieg Larsson	<i>Mannen die vrouwen haten</i>	3.90	5.81
238	Elizabeth Haynes	<i>Waarheen je ook vlucht</i>	3.89	6.01

Rank	Author	Title	Literary quality	General quality
239	Elena Forbes	<i>Sterf met mij</i>	3.89	5.73
240	Karin Slaughter	<i>Genadeloos</i>	3.89	5.71
241	Sophie Hannah	<i>Kleine meid</i>	3.88	5.49
242	David Baldacci	<i>Niets dan de waarheid</i>	3.88	5.95
243	Joy Fielding	<i>Roerloos</i>	3.88	5.78
244	Mo Hayder	<i>Diep</i>	3.87	5.76
245	Loes den Hollander	<i>Glansrol</i>	3.87	5.75
246	Tess Gerritsen	<i>Koud hart</i>	3.86	5.94
247	Camilla Läckberg	<i>Isprinses</i>	3.86	5.70
248	Tess Gerritsen	<i>Sneeuwval</i>	3.85	5.70
249	Mo Hayder	<i>Rot</i>	3.85	5.94
250	Harlan Coben	<i>Verloren</i>	3.85	5.87
251	Lynn Austin	<i>Eindelijk thuis</i>	3.85	5.75
252	Marion Pauw	<i>Drift</i>	3.85	5.29
253	Loes den Hollander	<i>Uitglijder</i>	3.85	5.63
254	John Grisham	<i>De bekentenis</i>	3.85	5.64
255	Jane Casey	<i>Spoorloos</i>	3.84	5.60
256	Jan Terlouw / Sanne Terlouw	<i>Hellehonden</i>	3.84	5.15
257	Loes den Hollander	<i>Wisselgeld</i>	3.84	5.45
258	Patricia Cornwell	<i>Rood waas</i>	3.83	5.66
259	David Baldacci	<i>In het geheim</i>	3.83	6.02
260	John Grisham	<i>De aanklacht</i>	3.83	5.69
261	David Baldacci	<i>De provocatie</i>	3.83	5.91
262	J.K. Rowling	<i>Harry Potter en de Relieken van de Dood</i>	3.83	6.09
263	David Baldacci	<i>Verlos ons van het kwaad</i>	3.82	5.69
264	Tess Gerritsen	<i>Verdwin</i>	3.82	5.96
265	Simone van der Vlugt	<i>Op klaarlichte dag</i>	3.81	5.22
266	Suzanne Vermeer	<i>De vlucht</i>	3.81	5.28
267	Elizabeth Gilbert	<i>Toewijding</i>	3.80	4.93
268	Harlan Coben	<i>Verzoeking</i>	3.79	5.78
269	James Worthy	<i>James Worthy</i>	3.79	4.17
270	David Baldacci	<i>De zesde man</i>	3.79	5.73
271	David Baldacci	<i>Geniaal geheim</i>	3.77	5.94
272	Jussi Adler-Olsen	<i>De vrouw in de kooi</i>	3.76	5.85
273	Loes den Hollander	<i>Broeinest</i>	3.76	5.62
274	Kim Moelands	<i>Weerloos</i>	3.75	5.54
275	Frederick Forsyth	<i>De cobra</i>	3.75	5.75
276	Karen Rose	<i>Moord voor mij</i>	3.75	6.05
277	Loes den Hollander	<i>Driftleven</i>	3.75	5.45
278	Karin Slaughter	<i>Gevallen</i>	3.74	5.63
279	Loes den Hollander	<i>Dwaalspoor</i>	3.74	5.38
280	Erica James	<i>Schaduwleven</i>	3.74	5.51
281	Cody McFadyen	<i>Tijd om te sterven</i>	3.74	6.17

Rank	Author	Title	Literary quality	General quality
282	Tess Gerritsen	<i>De Mefisto Club</i>	3.73	5.78
283	Mo Hayder	<i>Huid</i>	3.73	5.75
284	Karin Slaughter	<i>Verbroken</i>	3.73	5.57
285	Allan Folsom	<i>Dag van ontmaskering</i>	3.72	5.91
286	Erica James	<i>Zussen voor altijd</i>	3.72	5.45
287	Esther Verhoef	<i>Déjà vu</i>	3.72	5.09
288	Saskia Noort	<i>Afgunst</i>	3.72	5.09
289	Clive Cussler / Jack Du Brul	<i>Dodenschip</i>	3.71	5.96
290	Loes den Hollander	<i>Naaktportret</i>	3.71	5.69
291	David Baldacci	<i>Familieverraad</i>	3.71	5.91
292	David Baldacci	<i>De rechtvaardigen</i>	3.71	5.80
293	Jens Lapidus	<i>Snel geld</i>	3.70	5.20
294	David Baldacci	<i>Onschuldig</i>	3.70	5.79
295	James Patterson / Liza Marklund	<i>Partnerruil</i>	3.70	5.56
296	Judith Visser	<i>Stuk</i>	3.68	5.31
297	Marion Pauw	<i>Villa Serena</i>	3.68	5.18
298	Esther Verhoef	<i>Close-up</i>	3.67	5.19
299	Patricia Cornwell	<i>Mortuarium</i>	3.66	5.71
300	Jean M. Auel	<i>Het lied van de grotten</i>	3.66	5.22
301	Wilbur Smith	<i>Op volle zee</i>	3.66	5.55
302	Dan Brown	<i>Het verloren symbool</i>	3.65	5.38
303	James Patterson	<i>De affaire</i>	3.64	5.73
304	Saskia Noort	<i>Koorts</i>	3.63	4.91
305	Clive Cussler	<i>Medusa</i>	3.62	5.84
306	John Grisham	<i>Het proces</i>	3.61	5.50
307	Robert Ludlum / Kyle Mills	<i>Het Ares akkoord</i>	3.61	5.78
308	Suzanne Collins	<i>Vlammen</i>	3.61	6.03
309	Monika Peetz	<i>De dinsdagvrouwen</i>	3.59	4.87
310	David Baldacci	<i>Rechteloos</i>	3.59	5.69
311	Clive Cussler / Dirk Cussler	<i>Wassende maan</i>	3.58	5.83
312	Tom Clancy	<i>In het vizier</i>	3.58	5.96
313	James Patterson / Michael Ledwige	<i>Hitte</i>	3.54	5.63
314	Erica James	<i>De kleine dingen</i>	3.54	5.56
315	Naima El Bezaz	<i>Vinexvrouwen</i>	3.54	4.53
316	Lee Child	<i>De affaire</i>	3.53	5.97
317	Patricia Cornwell	<i>De Scarpetta factor</i>	3.53	5.48
318	Andy McNab	<i>Oorlogswond</i>	3.53*	5.53*
319	Carry Slee	<i>Bangkok boy</i>	3.52	4.81
320	Elizabeth Gilbert	<i>Eten, bidden, beminnen</i>	3.51	4.58
321	Kluun	<i>Haantjes</i>	3.50	4.31
322	Andy McNab	<i>Onbreekbare eenheid</i>	3.50*	5.56*

Rank	Author	Title	Literary quality	General quality
323	Karen Kingsbury	<i>Laatste dans</i>	3.50	5.37
324	Tom Clancy / Peter Telp	<i>De ogen van de vijand</i>	3.49	5.84
325	Emily Barr	<i>Klasgenoten</i>	3.49	5.39
326	Suzanne Collins	<i>Spotgaaï</i>	3.48	5.89
327	Karen Kingsbury	<i>Nooit te laat</i>	3.47	5.52
328	Suzanne Vermeer	<i>Bella Italia</i>	3.46	4.75
329	Suzanne Collins	<i>De hongerspelen</i>	3.46	5.86
330	Marion Pauw	<i>Jetset</i>	3.46	4.81
331	Fiona Neill	<i>Vriendschap, liefde en andere stommiteiten</i>	3.45*	5.03*
332	Maeve Binchy	<i>Hart & Ziel</i>	3.43	5.88
333	Christopher Paolini	<i>Erfenis</i>	3.42	4.60
334	Saskia Noort	<i>De eetclub</i>	3.42	5.43
335	Nora Roberts	<i>Eerbetoon</i>	3.41	5.61
336	Santa Montefiore	<i>Fairfield park</i>	3.40	5.53
337	Tom Clancy	<i>Op leven en dood</i>	3.38	5.62
338	Suzanne Vermeer	<i>De suite</i>	3.37	4.90
339	Clive Cussler / Dirk Cussler	<i>Duivelsadem</i>	3.37	5.91
340	Santa Montefiore	<i>In de schaduw van het palazzo</i>	3.37	5.36
341	Heleen van Royen	<i>Sabine †</i>	3.36*	4.88
342	Suzanne Vermeer	<i>Bon Bini beach</i>	3.35	5.05
343	Saskia Noort	<i>De verbouwing</i>	3.34	4.64
344	Steve Mosby	<i>50/50 Moorden</i>	3.34	5.67
345	Maeve Binchy	<i>En toen kwam Frankie</i>	3.33	5.39
346	Linda van Rijn	<i>Last minute</i>	3.32	5.18
347	Lee Child	<i>Tegenspel</i>	3.31	5.95
348	Julia Burgers-Drost	<i>Tussen hart en verstand</i>	3.29*	5.35*
349	Linda van Rijn	<i>Blue Curacao</i>	3.26	5.01
350	Merel van Groningen	<i>Misleid</i>	3.26	5.14
351	Lee Child	<i>61 Uur</i>	3.23	5.92
352	Santa Montefiore	<i>De affaire</i>	3.21	5.22
353	Santa Montefiore	<i>Villa Magdalena</i>	3.20	5.13
354	Heleen van Royen	<i>De mannentester</i>	3.18	4.32
355	Jennifer Weiner	<i>Sommige meisjes</i>	3.14*	5.41*
356	Van Sambeek	<i>Koninginnenrit</i>	3.12	4.94
357	Suzanne Vermeer	<i>Cruise</i>	3.11	4.38
358	Suzanne Vermeer	<i>Zwarte piste</i>	3.10	4.68
359	Mart Smeets	<i>Een koud kunstje †</i>	3.09	5.20
360	Suzanne Vermeer	<i>Après-ski</i>	3.07	4.49
361	Mart Smeets	<i>De afrekening</i>	3.06	4.96
362	Liza van Sambeek	<i>Bloed zaad en tranen</i>	3.05	4.75
363	Madeleine Wickham	<i>Het zwemfeestje</i>	2.98	5.15
364	Gerda van Wageningen	<i>In de schemering</i>	2.98*	5.15*

Rank	Author	Title	Literary quality	General quality
365	Danielle Steel	<i>Door dik en dun</i>	2.97	5.32
366	Danielle Steel	<i>De weg van het hart</i>	2.96	5.17
367	Candace Bushnell	<i>1 Fifth Avenue</i>	2.94	4.66
368	Madeleine Wickham	<i>Zoete tranen</i>	2.94	5.23
369	Appie Baantjer / Simon de Waal	<i>Een mes in de rug</i>	2.92	5.32
370	Cathy Kelly	<i>Eens in je leven</i>	2.91	5.35
371	J. Fallon	<i>Op de man af</i>	2.91	5.04
372	Clemens Wisse	<i>De jonge boerin van Madezicht</i>	2.90*	5.00*
373	Appie Baantjer / Simon de Waal	<i>Een schot in de roos</i>	2.88	5.22
374	Madeleine Wickham	<i>De vraagprijs</i>	2.88	5.30
375	Madeleine Wickham	<i>De cocktailclub</i>	2.87	5.12
376	Appie Baantjer / Simon de Waal	<i>Een rat in de val</i>	2.84	5.38
377	Mariette Middelbeek	<i>Turbulentie</i>	2.83	5.34
378	Hetty Luiten	<i>Je blijft altijd welkom</i>	2.83*	5.54*
379	Chantal van Gastel	<i>Geknipt voor jou</i>	2.83	5.66
380	Appie Baantjer / Simon de Waal	<i>Een lijk in de kast</i>	2.79	5.07
381	Jill Mansell	<i>Vlinders voor altijd</i>	2.78	5.34
382	Mariette Middelbeek	<i>Single en sexy</i>	2.74	5.23
383	Chantal van Gastel	<i>Zwaar verliefd!</i>	2.70	5.41
384	Jill Mansell	<i>De smaak te pakken</i>	2.70	5.39
385	Sabine van den Eynden / Emile Proper	<i>Gooische vrouwen</i>	2.69	4.49
386	Appie Baantjer / Simon de Waal	<i>Een dief in de nacht</i>	2.66	5.07
387	Jill Mansell	<i>Scherven brengen geluk</i>	2.65	5.26
388	E.L. James	<i>Vijftig tinten vrij</i>	2.64	4.65
389	Sophie Kinsella	<i>Wat spook jij uit?</i>	2.63	5.62
390	E.L. James	<i>Vijftig tinten donkerder</i>	2.60	4.51
391	Jill Mansell	<i>Eenmaal andermaal verliefd</i>	2.59	5.29
392	Jill Mansell	<i>Drie is te veel</i>	2.58	5.04
393	Katie Fforde	<i>Trouwplannen</i>	2.56	5.33
394	Lauren Weisberger	<i>Champagne in Chateau Marmont</i>	2.52	5.03
395	Jill Mansell	<i>Versier me dan</i>	2.50	5.17
396	Sophie Kinsella	<i>Ken je me nog?</i>	2.47	5.35
397	Sophie Kinsella	<i>Mini shopaholic</i>	2.46	4.84
398	Lauren Weisberger	<i>Chanel chic</i>	2.45	4.71
399	Sophie Kinsella	<i>Mag ik je nummer even?</i>	2.44	5.52
400	Sophie Kinsella	<i>Shopaholic en baby</i>	2.37	4.89
401	E.L. James	<i>Vijftig tinten grijs</i>	2.12	3.54

Appendix 3: The Website

At <https://literaryquality.huygens.knaw.nl/> a link is presented to the companion website. The website contains:

- Colour versions of the charts in this book;
- Additional figures, some of which are interactive;
- Explanations of the extra figures;
- Csv files and an R Package with the data from The National Reader Survey, the metadata for the 401 novels in the research corpus and a number of linguistic data of the novels, all with appropriate documentation;
- Other information about the research and about this book.

Bibliography

All hyperlinks were last accessed in February 2023.

- Allison, Sarah. *Reductive Reading: A Syntax of Victorian Moralising*. Johns Hopkins University Press, 2018.
- Anthony, Laurence. *AntConc*. Version 3.5.8., Tokyo: Waseda University, 2019, <https://www.laurenceanthony.net/software.html>.
- Archer, Jodie, and Matthew L. Jockers. *The Bestseller Code*. Penguin Books, 2016.
- Bassnett, Susan. *Translation*. Routledge, 2014.
- Bawarshi, Anis S. 'The Genre Function.' *College English*, vol. 62, no. 3, 2002, pp. 335–360, doi:10.2307/378935.
- Bawarshi, Anis S. *Genre and the Invention of the Writer: Reconsidering the Place of Invention in Composition*. Utah State University Press, 2003.
- Bellos, David. *Is That a Fish in your Ear?: The Amazing Adventure of Translation*. Penguin Books, 2012.
- Bertens, Hans. *Literary Theory: The Basics*. 2nd ed., Taylor & Francis, 2008.
- Beyers, Ton, et al. *Nederlandse kunst in de wereld: Literatuur, architectuur en beeldende kunst 1980-2013* [*Dutch Art in the World: Literature, Architecture and Visual Arts 1980-2013*, not available in English]. Vantilt, 2015.
- Blei, David M. 'Probabilistic Topic Models'. *Communications of the ACM* vol. 55, no. 4, 2012, pp. 77–84, doi:10.1145/2133806.2133826.
- Bloom, Paul. *How Pleasure Works: The New Science of Why We Like What We Like*. Norton, 2010.
- Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary History*. University of Michigan Press, 2018, doi:10.3998/mpub.8784777.
- Bouma, Gosse, et al. 'Alpino: Wide-Coverage Computational Analysis of Dutch.' *Language and Computers*, vol. 37, no. 1, 2002, pp. 45–59, <https://www.let.rug.nl/vannoord/papers/alpino.pdf>.
- Bourdieu, Pierre. *Distinction: A Social Critique of the Judgement of Taste*. 1984. Translated by Richard Nice. Routledge, 2010.
- Bourdieu, Pierre. *The Field of Cultural Production: Essays on Art and Literature*, edited by Randal Johnson. Polity, 1993.
- Brems, Elke, et al., editors. *Doing Double Dutch: The International Circulation of Literature from the Low Countries*. Leuven University Press, 2017.
- Burrows, John. "'Delta": A Measure of Stylistic Difference and a Guide to Likely Authorship.' *Literary and Linguistic Computing*, no. 17, 2002, pp. 267–287, doi:10.1093/lc/17.3.267.
- Calvo Tello, José. *The Novel in the Spanish Silver Age. A Digital Analysis of Genre Using Machine Learning*. Bielefeld University Press, 2021, doi:10.14361/9783839459256.

- 'Computational Literary Studies Infrastructure.' *European Commission Community Research and Development Information Service (CORDIS)*, 28 April 2022, <https://cordis.europa.eu/project/id/101004984>.
- Computational Stylistics Group, <https://computationalstylistics.github.io/>.
- Craig, Hugh, and Brett Greatley-Hirsch. *Style, Computers, and Early Modern Drama: Beyond Authorship*. Cambridge University Press, 2017.
- Culpeper, Jonathan. 'Keywords and Characterization. An Analysis of Six Characters in Romeo and Juliet.' David L. Hoover et al. *Digital literary studies. Corpus approaches to poetry, prose, and drama*. Routledge, 2014, pp. 9–34.
- Deneire, Tom, et al. 'A Game with Words: Rhetorical Citizenship and Game Theory'. *Contemporary Rhetorical Citizenship*, edited by Christian Kock and Lisa Villadsen. Leiden University Press, 2014, pp. 323–340.
- Dera, Jeroen. 'De lezende leraar: Literatuuronderwijs in Nederland(s) als onderzoeksobject' ['The Reading Teacher: Dutch Literary Education in the Netherlands as an Object of Research,' not available in English]. *Tijdschrift voor Nederlandse Taal- en Letterkunde*, vol. 134, no. 2, 2018, pp. 146–170.
- Dera, Jeroen. *De praktijk van de leeslijst: Een onderzoek naar de inhoud en waardering van literatuurlijsten voor het schoolvak Nederlands op havo en vwo* [*The Practice of the Reading List: A Study of the Content and Appreciation of Reading Lists for Dutch as School Subject in Senior General Secondary Education and Pre-University Education*, not available in English]. Stichting Lezen, 2019, https://www.lezen.nl/sites/default/files/praktijk_van_de_leeslijst_digitaal.pdf.
- 'Distant Reading for European Literary History.' European Cooperation in Science and Technology (COST Action CA16204)", <https://www.cost.eu/actions/CA16204/>.
- Dixon, Peter, et al. 'Judging a Book by its Cover.' *Scientific Study of Literature*, vol. 5, no. 1, 2015, pp. 23–48, doi:10.1075/ssol.5.1.02dix.
- Dorrestein, Renate. 'Ouverture.' *De zondagmiddagauto* [*The Sunday Afternoon Car*, not available in English. B for Books Publishing, 2012, pp. 7–26.
- Dorrestein, Renate. *Onder de motorkap van het schrijverschap* [*Under the Bonnet of Authorship*, not available in English]. Em. Querido's Uitgeverijen, 2017.
- Dutton, Denis. *The Art Instinct: Beauty, Pleasure, and Human Evolution*. Oxford University Press, 2009.
- Eder, Maciej. 'Does Size Matter? Authorship Attribution, Small Samples, Big Problem.' *Digital Scholarship in the Humanities*, vol. 30, no. 2, 2015, pp. 167–182, doi:10.1093/lc/fqt066.
- Eder, Maciej, et al. 'Stylometry with R: A Package for Computational Text Analysis'. *The R Journal*, vol. 8, no. 1, 2016, pp. 107–121, <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html>.
- European Commission, Directorate-General for Education, Youth, Sport and Culture, *Successful PISA Stories in the EU: How Some Member States Have Been Able to Improve*

- Their Performance over Time*, final report, Publications Office of the European Union, 2022, <https://data.europa.eu/doi/10.2766/429517>.
- Eve, Martin Paul. *Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell's Cloud Atlas*. Stanford University Press, 2019.
- Felski, Rita. *Uses of Literature*. Blackwell Press, 2008.
- Fortuin, Arjen. 'Mogen wij nog haantjes zijn? Tweestrijd van de moderne man is het gemeenschappelijke thema van Kluun en Koch' ['Are We Still Allowed to Be Alpha Males? The Inner Turmoil of the Modern Man is the Communal Topic of Kluun and Koch', not available in English]. *nrc-next*, 28 January 2011, pp. 4–5, <https://www.nrc.nl/nieuws/2011/01/28/mogen-wij-nog-haantjes-zijn-11991928-a1367099>.
- Franssen, Thomas. *How Books Travel: Translation Flows and Practices of Dutch Acquiring Editors and New York Literary Scouts, 1980-2009*. 2014. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245/1.446961>.
- Hennion, Antoine. 'Those Things that Hold Us Together: Taste and Sociology.' *Cultural Sociology*, vol. 1, no. 1, 2007, pp. 97–114, doi:10.1177/1749975507073923.
- Herrmann, Berenike J., et al. 'Revisiting Style, a Key Concept in Literary Studies.' *Journal of Literary Theory*, no. 9, 2015, pp. 25–52.
- Hoover, David L. 'Testing Burrows's Delta.' *Literary and Linguistic Computing*, vol. 19, no. 4, 2004, pp. 453–475, doi:10.1093/lc/19.4.453.
- Hoover, David L. *Modes of Composition and the Durability of Style in Literature*. Routledge, 2021.
- Janssen, Elisa. 'We hebben wel heel veel mannen, laten we nu eens een vrouw nemen': Een verkennend onderzoek naar een genderkloof in het literatuuronderwijs in de bovenbouw van havo en vwo. ['We Have Rather a Lot of Men, Let's Add a Woman for a Change: An Exploratory Study into a Gender Gap in Literature Education in Advanced Stage Senior General Secondary Education and Pre-University Education', not available in English]. 2019. University of Amsterdam, master's thesis. *University of Amsterdam Digital Academic Repository*, https://scripties.uba.uva.nl/search?id=record_23761.
- Jautze, Kim, et al. 'From High Heels to Weed Attics: A Syntactic Investigation of Chick Lit and Literature'. *Proceedings of the Second Workshop on Computational Linguistics for Literature*, Atlanta, 14 June 2013, <https://aclanthology.org/W13-1410/>.
- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013.
- Journal of Computational Literary Studies* (JCLS), <https://jcls.io/>.
- Kestemont, Mike, et al. 'Collaborative Authorship in the Twelfth Century: A Stylometric Study of Hildegard of Bingen and Guibert of Gembloux'. *Digital Scholarship in the Humanities*, vol. 30, no. 2, 2015, pp. 199–224. doi:10.1093/lc/fqt063.
- Ketzan, Erik and Christof Schöch. 'What Changed When Andy Weir's *The Martian* Got Edited?' *Digital Humanities 2017 Conference Abstracts*, edited by Rhian Lewis et al. McGill University & Université de Montréal, 2017, pp. 285–288.

- Ketzan, Erik and Christof Schöch. 'Classifying and Contextualizing Edits in Variants with Coletto: Three Versions of Andy Weir's *The Martian*'. *Digital Humanities Quarterly (DHQ)* vol. 15, no. 4, 2021, <http://digitalhumanities.org:8081/dhq/vol/15/4/000579/000579.html>.
- Koolen, Corina. *Reading Beyond the Female: The Relationship Between Perception of Author Gender and Literary Quality*. 2018. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37fice7>.
- Koolen, Corina, et al. 'Literary Quality in the Eye of the Dutch Reader: The National Reader Survey.' *Poetics*, vol. 79, no. 101439, 2020, pp. 1–13, doi:10.1016/j.poetic.2020.101439.
- Lauwers, Jeroen, et al. 'The Question of Literary Quality: A Rhetorical and Game-Theoretical Approach'. *Scientific Study of Literature*, vol. 5, no. 1, 2015, pp. 3–22, doi: 10.1075/ssol.5.1.01lau.
- Lensink, Saskia, et al. 'Hidden Indicators: Modal Particles in Modern Dutch Fiction and Fiction Translated from English, A Quantitative Approach.' Unpublished manuscript, https://tonvanderwouden.nl/index_files/papers/Lensink-Wouden-Dalen.pdf.
- 'Letterenfonds Vertalingendatabase' ['Dutch Foundation for Literature Translation Database']. *Dutch Foundation for Literature*, <https://www.letterenfonds.nl/nl/vertalingendatabase>.
- Lewandowska-Tomaszczyk, Barbara. 'Explicit and Tacit: An Interplay of the Quantitative and Qualitative Approaches to Translation'. *Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research*, edited by Michael P. Oakes en Meng Ji. John Benjamins Publishing Company, 2012, pp. 3–34.
- McCarthy, Rachel and James O'Sullivan. 'Who wrote *Wuthering Heights*?' *Digital Scholarship in the Humanities*, vol. 36, no. 2, 2021, pp. 383–391, doi:10.1093/llc/fqaa031.
- 'Novel Perceptions: Towards an Inclusive Canon Project Website.' *University of Wolverhampton*, <http://novel-perceptions.thememorynetwork.com/>.
- Oakes, Michael P., and Meng Ji, editors. *Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research*, Studies in Corpus Linguistics 51. John Benjamins Publishing Company, 2012.
- 'Organisation for Economic Co-operation and Development (OECD).' *Programme for International Student Assessment (PISA)*, <https://www.oecd.org/pisa/>.
- Pearl, Lisa et al. 'The Character in the Letter: Epistolary Attribution in Samuel Richardson's *Clarissa*'. *Digital Scholarship in the Humanities*, vol. 32, no. 2, 2017, pp. 355–376, doi:10.1093/llc/fqw007.
- Van Peer, Willie, editor. *The Quality of Literature: Linguistic Studies in Literary Evaluation*. John Benjamins Publishing Company, 2008.
- Peters, Arjan. 'De nieuwe Kluun mag er wezen: Hoogstaand amusement in Haantjes van Kluun. Of: hoe Stijn van Diepen goud geld dacht te verdienen aan de Gay Games in Amsterdam'. ['The Latest Kluun Deserves Praise: Elevated Entertainment in *Alpha Males* by Kluun. Or: How Stijn van Diepen Schemed to Get Filthy Rich by Commercially Exploiting

- the Amsterdam Gay Games', not available in English]. *De Volkskrant*, 23 January 2011, <https://www.volkskrant.nl/cultuur-media/de-nieuwe-kluun-mag-er-wezen~bof5a268/>.
- Piper, Andrew. *Enumerations: Data and Literary Study*. University of Chicago Press, 2018.
- Primorac, Antonija, et al. 'Distant Reading Two Decades On: Reflections on the Digital Turn in the Study of Literature.' To be published in *Digital Studies / Le champ numérique*, doi:10.16995/dscn.8855.
- Radway, Janice A. *Reading the Romance: Women, Patriarchy, and Popular Literature*. 1984. Updated ed. University of North Carolina Press, 1991.
- Riddell, Allen, and Karina van Dalen-Oskam, 'Readers and their Roles: Evidence from Readers of Contemporary Fiction in the Netherlands.' *PLoS ONE*, vol. 13, no. 7, 2018, pp. 1–12, doi:10.1371/journal.pone.0201157.
- 'The Riddle of Literary Quality Project Website.' *Huygens Institute, Royal Netherlands Academy of Arts and Sciences*, https://literaryquality.huygens.knaw.nl/?page_id=179.
- Rybicki, Jan. 'The Great Mystery of the (Almost) Invisible Translator: Stylometry in Translation.' *Quantitative Methods in Corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research*, edited by Michael P. Oakes en Meng Ji. John Benjamins Publishing Company, 2012, pp. 231–248.
- Rybicki, Jan, and Magda Heydel. 'The Stylistics and Stylometry of Collaborative Translation: Woolf's Night and Day in Polish'. *Literary and Linguistic Computing*, vol. 28, no. 4, 2013, pp. 708–717, doi:10.1093/lc/fqt027.
- Rybicki, Jan, et al. 'Collaborative Authorship: Conrad, Ford and Rolling Delta'. *Literary and Linguistic Computing*, vol. 29, no. 3, 2014, pp. 422–431, doi:10.1093/lc/fqu016.
- Salgaro, Massimo. 'How Literary Can Literariness Be? Methodological Problems in the Study of Foregrounding.' *Scientific Study of Literature*, vol. 5, no. 2, 2015, pp. 229–249, doi: 10.1075/ssol.5.2.06sal.
- Schwarz, Ori. 'Bending Forward, One Step Backward: On the Sociology of Tasting Techniques'. *Cultural Sociology*, vol. 7, no. 4, 2013, pp. 415–430, doi:10.1177/1749975513480957.
- Slob, Marjan. *Foute fantasieën: Kleine filosofie van de ontvankelijkheid*. [Bad Fantasies: Small Philosophy of Receptivity, not available in English]. 2007. Updated ed. Lemniscaat, 2017.
- Squires, Claire. *Marketing Literature: The Making of Contemporary Writing in Britain*. Palgrave Macmillan, 2009.
- Stukker, Ninke, and Arie Verhagen. *Stijl, taal en tekst: Stilistiek op taalkundige basis* [Style, Language and Text: Linguistics-Based Stylistics, not available in English]. Leiden University Press, 2019.
- Taylor, Helen. *Why Women Read Fiction: The Stories of Our Lives*. Oxford University Press, 2019.
- Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019.
- Van Cranenburgh, Andreas, and Corina Koolen. 'Identifying Literary Novels with Bigrams.' *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, Denver, June 2015, pp. 58–67, doi: 10.3115/v1/W15-0707.

- Van Cranenburgh, Andreas. *Rich Statistical Parsing and Literary Language*. 2016. University of Amsterdam, PhD dissertation. *University of Amsterdam Digital Academic Repository*, <https://hdl.handle.net/11245.1/cb936704-8215-4f47-9013-0d43d37fice7>.
- Van Cranenburgh, Andreas, and Rens Bod. 'A Data-Oriented Model of Literary Language.' *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, Valencia, 3–7 April 2017, pp. 1228–1238, <https://aclanthology.org/E17-1115/>.
- Van Cranenburgh, Andreas. 'Cliché Expressions in Literary and Genre Novels.' *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Santa Fe, 25 August 2018, pp. 34–43, <https://aclanthology.org/W18-4504/>.
- Van Cranenburgh, Andreas, et al. 'Vector Space Explorations of Literary Language.' *Language Resources and Evaluation*, vol. 53, 2019, pp. 625–650, doi:10.1007/s10579-018-09442-4.
- Van Cranenburgh, Andreas, and Corina Koolen. 'Results of a Single Blind Literary Taste Test with Short Anonymized Novel Fragments.' *Proceedings of The 4th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Barcelona, 12 December 2020, pp. 121–126, <https://aclanthology.org/2020.latechclfl-1.14.pdf>.
- Van Dalen-Oskam, Karina, and Joris van Zundert. 'Delta for Middle Dutch: Author and Copyist Distinction in *Walewein*'. *Literary and Linguistic Computing*, vol. 22, no. 3, 2007, pp. 345–362, doi:10.1093/lc/fqm012.
- Van Dalen-Oskam, Karina. 'Names in Novels: An Experiment in Computational Stylistics.' *LLC: The Journal of Digital Scholarship in the Humanities*, vol. 28, no. 2, 2013, pp. 359–370, doi: 10.1093/lc/fqs007.
- Van Dalen-Oskam, Karina. 'Epistolary Voices: The Case of Elisabeth Wolff and Agatha Deken'. *LLC: The Journal of Digital Scholarship in the Humanities*, no. 29, vol. 3, 2014, pp. 443–451, doi: 10.1093/lc/fqu023.
- Van Dalen-Oskam, Karina. 'A Literary Rat-Race.' *Digital Humanities Conference 2016*, Kraków, 11–16 July 2016. Conference Abstract. <https://www.dh2016.adho.org/abstracts/267>.
- Van Dalen-Oskam, Karina. *Het raadsel literatuur: Is literaire kwaliteit meetbaar? [The Riddle of Literature: Can We Measure Literary Quality?]* Amsterdam University Press, 2021, <https://karinavdo.github.io/RaadselLiteratuur/>.
- Van der Deijl, Lucas, et al. 'Mapping the Demographic Landscape of Characters in Recent Dutch Prose: A Quantitative Approach to Literary Representation'. *Journal of Dutch Literature*, vol. 7, no. 1, April 2016, <https://www.journalofdutchliterature.org/index.php/jdl/article/view/103>.
- Van Dijk, Yra. *Afgrond zonder vangnet: Liefde en geweld in het werk van Arnon Grunberg [Abyss Without a Safety Net: Love and Violence in the Work of Arnon Grunberg]*, not available in English]. Nijgh and Van Ditmar, 2018.
- Van Dijk, Yra, et al. *Omdat lezen loont. Op naar effectief leesonderwijs in Nederland [Because Reading is Beneficial. Towards Effective Reading Pedagogy in the Netherlands]*, not available in English]. Pica, 2022.

- Van het Reve, Karel. 'Literatuurwetenschap: het raadsel der onleesbaarheid' ['Literary Theory: The Riddle of Illegibility,' not available in English]. Johan Huizinga Lecture, 8 December 1978, Leiden, https://www.dbnl.org/tekst/reveo03lite02_01/index.php.
- Van Peer, Willie, editor. *The Quality of Literature: Linguistic Studies in Literary Evaluation*. John Benjamins Publishing Company, 2008.
- Van Rossum, Lisanne M. *Towards Literary Narrative Perspective: The Relationship Between Narrative Perspective and Perceived Literariness*. 2020. Utrecht University, research master's thesis. *Utrecht University Repository*, <https://studenttheses.uu.nl/handle/20.500.12932/38257>.
- Van Rossum, Lisanne M., et al. 'I Catching: Computationally Operationalising Narrative Perspective for Stylometric Analysis.' *DH Benelux Journal*, vol. 3, 2021, pp. 1–12, doi:10.5281/zenodo.3855653.
- Van Wingerden, Wouter, and Pepijn Hendriks. 'Dat hoor je mij niet zeggen!' ['You Will Not Catch Me Saying That!,' not available in English]. Thomas Rap, 2015.
- Venuti, Lawrence. *The Translator's Invisibility: A History of Translation*. 1995. 2nd ed., Routledge, 2008.
- Verstraeten, Pieter. 'Na de race: Sport en literatuur rond 1900' ['After the Race: Sport and Literature Around 1900']. *Uit de marge: Kanttekeningen bij de cultuurhistorische canon* [*From the Margin: Side Notes to the Socio-Historical Canon*, not available in English], edited by Lizet Duyvendak and Jan Oosterholt. *Literatoren*, 2018, pp. 207–2013.
- Von Heydebrand, Renate, and Simone Winko. *Einführung in die Wertung von Literatur: Systematik – Geschichte – Legitimation* [*Introduction to the Evaluation of Literature: Systematics – History – Legitimacy*, not available in English]. Ferdinand Schöningh, 1996.

Tables

Table 2.1:	Top 10 most read of the 401 books	44
Table 2.2:	Top 10 most literary of the 401 books	45
Table 2.3:	Top 10 least literary of the 401 books	46
Table 2.4:	Top 10 books participants had not read but had an opinion about	48
Table 2.5:	Mean scores for literary quality and general quality per category	49
Table 3.1:	Five novels by Henning Mankell	77
Table 3.2:	Categories and number of books originally in Dutch and translated into Dutch	82
Table 3.3:	Categories in Dutch and translated with mean score for literary quality	82
Table 3.4:	Categories of Literary novel and Suspense in Dutch and translated with mean score for literary quality	82
Table 6.1:	The ten novels by Dutch-language female and male authors with the highest scores for literary quality	171
Table 6.2:	Personal pronouns more prominent in novels with the highest scores for literary quality compared to those with the lowest scores and vice versa	175
Table 6.3:	Verb tense more prominent in novels with the highest scores for literary quality compared to those with the lowest scores and vice versa	175

Figures

Figure 2.1: Number of books respondents indicated to have read per year	39
Figure 2.2: Number of books read (left) and number of books rated (right)	42
Figure 2.3: Number of books read per category	44
Figure 2.4: Mean scores for all 401 books	47
Figure 2.5: For all 401 books the mean score for literary quality per category	50
Figure 2.6: For all 401 books the mean score for general quality per category	50
Figure 2.7: Ratings of Tatiana de Rosnay, <i>Haar naam was Sarah</i> (<i>Sarah's Key</i>)	51
Figure 2.8: Number of books read, by gender author and gender reader	58
Figure 2.9: Number of books rated, by gender author and gender reader	59
Figure 3.1: The ten novels with the highest and lowest scores for literary quality, Cluster Analysis (100 most frequent words)	65
Figure 3.2: Ratings for the literary quality of E.L. James, <i>Vijftig tinten</i> -trilogie (<i>Fifty Shades</i> Trilogy)	67
Figure 3.3: E.L. James, <i>Vijftig tinten</i> -trilogie (<i>Fifty Shades</i> Trilogy) and Romance, PCA (1000 most frequent words)	71
Figure 3.4: E.L. James, <i>Vijftig tinten</i> -trilogie (<i>Fifty Shades</i> Trilogy) and Suspense written by female authors, PCA (1000 most frequent words)	72
Figure 3.5: Translated Literary novels written by male authors and three Suspense novels by Mankell, PCA (1000 most frequent words)	79
Figure 3.6: Literary novels by female authors written in Dutch and translated from English, PCA (1000 most frequent words)	87
Figure 3.7: The conjunctions <i>en</i> (and), <i>toen</i> (then), <i>dat</i> (that), <i>voordat</i> (before), <i>terwijl</i> (while), and <i>hoewel</i> (although) in Literary novels by female authors written in Dutch and translated from English, PCA	88
Figure 4.1: Ratings of Nicole Krauss, <i>Het grote huis</i> (<i>Great House</i>)	97
Figure 4.2: Ratings of Elizabeth Gilbert, <i>Eten, bidden, beminnen</i> (<i>Eat, Pray, Love</i>)	98
Figure 4.3: Translated Literary novels written by female authors, PCA (1000 most frequent words)	102
Figure 4.4: Ratings of Emma Donoghue, <i>Kamer</i> (<i>Room</i>)	105
Figure 4.5: Ratings of Heleen van Royen, <i>De mannentester</i> (<i>*The Tester of Men</i>)	109

Figure 4.6: Dutch Literary novels written by female authors, PCA (1000 most frequent words)	110
Figure 4.7: Dutch Literary novels written by female authors, Cluster Analysis (1000 most frequent words)	115
Figure 5.1: Ratings of Julian Barnes, <i>Alsof het voorbij is</i> (<i>The Sense of an Ending</i>)	126
Figure 5.2: Translated Literary novels written by male authors, PCA (1000 most frequent words)	130
Figure 5.3: Ratings of Erwin Mortier, <i>Gestameld liedboek</i> (<i>Stammered Songbook</i>)	135
Figure 5.4: Ratings of Mart Smeets, <i>De afrekening</i> (<i>*The Reckoning</i>)	136
Figure 5.5: Dutch Literary novels written by male authors, PCA (1000 most frequent words)	139
Figure 5.6: Ratings of Kluun, <i>Haantjes</i> (<i>*Alpha Males</i>)	144
Figure 5.7: Ratings of Herman Koch, <i>Zomerhuis met zwembad</i> (<i>Summerhouse with Swimming Pool</i>)	144
Figure 5.8: Kluun, <i>Haantjes</i> (<i>*Alpha Males</i>), versus Koch, <i>Zomerhuis met zwembad</i> (<i>Summerhouse with Swimming Pool</i>), Cluster Analysis (1000 most frequent words)	147
Figure 5.9: Grunberg, <i>Huid en haar</i> (<i>*Hair and Hide</i>) en <i>De man zonder ziekte</i> (<i>*The Man without Illness</i>), versus Kinsella, <i>Shopaholic en baby</i> (<i>Shopaholic and Baby</i>) en <i>Ken je me nog?</i> (<i>Remember me?</i>), Cluster Analysis (1000 most frequent words)	154
Figure 6.1: Mean sentence length for all 401 books	169
Figure 6.2: Sentence length variation for all 401 books	169
Figure 6.3: Literary quality top 10 of Dutch-language novels by female and male authors	171
Figure 6.4: Literary quality top 10 of Dutch-language novels by female and male authors, Cluster Analysis (1000 most frequent words)	172
Figure 6.5: Literary quality top 10 of Dutch-language novels by female and male authors, PCA (1000 most frequent words)	173
Figure 6.6: Mean sentence length (in words) of Dutch-language Literary novels	174

Acknowledgements

This book is an abbreviated and slightly reorganized version of the original Dutch book *Het raadsel literatuur. Is literaire kwaliteit meetbaar?* that was published by Amsterdam University Press in November 2021. It presents an overview and interpretation of the results of the project The Riddle of Literary Quality of which I was the principal investigator. The project was funded by the Computational Humanities Programme (2012–2016) of the Royal Netherlands Academy of Arts and Sciences (KNAW). I am also grateful for the continued support of the project by three subsequent directors of Huygens Institute – KNAW, where the research was mainly executed: Henk Wals, Lex Heerma van Voss, and Dirk van Miert.

Many people played a role in the project in one way or another. First and foremost, warmest thanks are due to the members of the core project team: Hayco de Jong, Kim Jautze, Corina Koolen, and Andreas van Cranenburgh. Rens Bod and Hanno Brand were closely involved in the proposal and co-supervised the project; Andrea Scharnhorst and Sally Wyatt guided the project in the KNAW eHumanities Group. Furthermore, the team owes a lot to the input of the international think tank that shared their advice generously and with genuine interest: David Hoover, Patrick Juola, Jelle Krol, Max Louwerse, Martin Reynaert, Thomas Vaessens, Arie Verhagen, and Jelle Zuidema.

The project was supported in different ways by many other people working at Huygens Institute and elsewhere: Peter Boot, Stijn Daenekindt, Bas Doppen, Maciej Eder, Gertjan Filarski, Thomas Franssen, Guido Gerritsen, Johan Kwantes, Saskia Lensink, Fernie Maas, Erica Nagelhout, Allen Riddell, Kyra Schiffers, Renske Siemens, and Ton van der Wouden. Also thanked are the students at the University of Amsterdam who wrote, supervised by Everdien Rietstap, Lisa Kuitert, and myself, their master's thesis about The Riddle of Literary Quality: Lili Burki, Nanko van Egmond, Nynke de Groot, Anne ter Hark, Sterre Houweling, Anna Krans, Kevin Smit, Sophie Visser, and Jasmijn Zondervan.

Bas Groes invited me to be part of a team at the University of Wolverhampton to replicate the Riddle research in the United Kingdom. Many thanks to Bas and to the other members of the Novel Perceptions team for this inspiring collaboration as well as for commenting on earlier versions of this book: Aidan Byrne, Elizabeth Dearnley, Peter Harvey, Adrian Leguina, Tom Mercer, Emad Mohamed, Marjolein Poortvliet, Lisanne van Rossum, and Demi-Mae Wilton.

For a first rough translation into English of many sections of the Dutch book I used DeepL, which I then shortened and edited. Lisanne van Rossum then edited my text, after which I gave the book a final edit. Joris van Zundert provided many of the figures in the book, a full-colour version of which can be found on the companion

website. I am grateful to Christof Schöch and the other, anonymous, peer reviewers for providing excellent and detailed comments that were very helpful. Thanks are also due to Katrien Depuydt, Maciej Eder, Bas Groes, Lianne van Rossum, Andreas van Cranenburgh, André van Dalen, and Joris van Zundert for their advice on specific issues. I thank Anja van Leusden at Amsterdam University Press for the pleasant and practical guidance of the book towards publication.

Special thanks go to Maciej Eder, for developing the litRiddle package for R in collaboration with Saskia Lensink and Joris van Zundert; to Lianne van Rossum, for her help in so many ways during the writing process; to Joris van Zundert, for our fruitful collaboration throughout the years and for building the companion website; to Katrien Depuydt, for her friendship and continuing support, and to André van Dalen, for everything.

Index

- Abdolah, Kader 133
De koning (The King) 133
*De kraai (*The Crow)* 133
- Adler-Olsen, Jussi 78, 80
*De bedrijfsterrorist (*The Corporate Terrorist)* 80
- Allende, Isabel 100, 168
Het eiland onder de zee (Island Beneath the Sea) 100, 168
Het negende schrift van Maya (Maya's Notebook) 100
- Allison, Sarah 28
- Amirrezvani, Anita 100
Dochter van Isfahan (The blood of flowers) 100
- Ammaniti, Niccolò 56, 59, 128, 165
Ik en jij (Me and You) 128, 165
Laat het feest beginnen (Let the Games Begin) 128
*Het laatste oudejaar van de mensheid (*The Last New Year's Eve of Humanity)* 128
Zo God het wil (As God Commands) 128
- AntConc 52, 55, 64, 85, 89, 101, 121, 130, 140, 174, 176, 190
- Anthony, Laurence 52, 55, 64, 85, 101, 130, 140, 174, 176
- Archer, Jodie 23-25, 162
- Auel, Jean M. 101, 166
Het lied van de grotten (The Land of Painted Caves) 101, 166
- Austen, Jane 22, 24
- Austin, Lynn 101
Eindelijk thuis (Until We Reach Home) 101
- Authorship attribution 19-22, 25, 28, 70
- Avallone, Silvia 81, 100
Staal (Swimming to Elba) 81, 100
- Baantjer, Appie 46, 56, 57, 59
*Een dief in de nacht (*A Thief in the Night)* 46
- Bakker, Gerbrand 132, 165, 171, 172
De omweg (The Detour) 132, 165, 171, 172
- Baldacci, David 43, 59, 125, 129
Die zomer (One Summer) 125, 129
- Barnes, Julian 45, 64-65, 124-127, 134, 146, 147, 153, 154, 165, 170, 182
Alsof het voorbij is (The Sense of an Ending) 45, 64, 124-127, 134, 146, 147, 153, 154, 165, 170, 182
- Barr, Emily 101
Klasgenoten (Out Of My Depth) 101
- Bassnett, Susan 83, 90, 91
- Bauer, Belinda 182
Snap 182
- Bawarshi, Anis 183-184
- Belgium 10, 38, 125, 189
- Bellos, David 83, 85
- Bernlef, J. 45, 56, 132, 171
*De een zijn dood (*One Man's Meat)* 132
*Geleende levens (*Borrowed Lives)* 45, 132, 171
- Bevers, Ton 10, 90
- Big Book Review 12, 37, 197
- Binchy, Maeve 58
- Binet, Laurent 128
HhhH (HHhH) 128
- Blake, Sarah 100, 101
De laatste brief (The Postmistress) 100, 101
- Bloom, Paul 17
- Blum, Jenna 100
Het familieportret (Those Who Save Us) 100
In tweestrijd (The Stormchasers) 100
- Bod, Rens 165
- Bode, Katherine 28
- Book week gift 36, 37, 138
- Booker Prize 45, 126, 182
- Bourdieu, Pierre 9, 10, 13, 14, 15, 16, 28, 182, 193, 194
- Boyne, John 81, 128
Het winterpaleis (The House Of Special Purpose) 81, 128
- Brems, Elke 10
- Brijs, Stefan 132
*Post voor mevrouw Bromley (*Mail for Miss Bromley)* 132
- Brokken, Jan 37
*Baltische zielen (*Baltic Souls)* 37
- Brontë, Branwell 21
- Brontë, Emily 21
Wuthering Heights 21
- Brontë sisters 114
- Brown, Dan 24, 59
Inferno 24
- Brown, Janelle 100
Alles wat wij wilden was alles (All We Ever Wanted Was Everything) 100
- Burrows, John 19, 20
- Burrows' Delta *see* Delta
- Bushnell, Candace 66, 182
1 Fifth Avenue (One Fifth Avenue) 66, 182
- Buwalda, Peter 44, 133
Bonita Avenue (Bonita Avenue) 44, 133
- Calvo Tello, José 28, 182-184
- Campert, Remco 133
*Dagboek van een poes (*Diary of a Cat)* 133
- Canon, the literary 12, 14, 25, 28, 188, 195
- Chicklit.nl* 34
- Child, Lee 47, 193
De affaire (The Affair) 193
6i Uur (6i Hours) 47
- Cleave, Chris 128
Kleine Bij (The Other Hand) 128

- Cluster analysis 63, 64, 70, 114, 116, 121, 122, 147, 148, 172, 190
- Coelho, Paulo 128
Aleph (Aleph) 128
De beschermengel (The Valkyries) 128
- Collins, Jacky 24
- Collins, Suzanne 52
- Computational Literary Studies Infrastructure (CLS Infra) 18
- Conrad, Joseph 21
Nostramo 21
- COST Action Distant Reading for European Literary History 18
- CPNB (Foundation 'Collective Promotion for the Dutch Book') 35, 36
- Cronin, Justin 125, 129
De oversteek (The Passage) 125, 129
- Cussler, Clive en Dirk Cussler 168
Duivelsadem (Arctic Drift) 168
- Damsma, Harm 131-132
- De Jong, Hayco 159
- Deken, Agatha 21
- De Loo, Tessa 37, 106
*De grote moeder (*The Great Mother)* 37, 106
*De meisjes van de suikerwerkfabriek (*The Girls from the Confectionery Factory)* 37
- Delta 19-21, 63, 64, 95, 114, 124, 190
- De Moor, Marente 89, 95, 106, 110, 112-115, 122, 137, 171
De Nederlandse maagd (The Dutch Maiden) 89, 106, 110, 112-115, 137, 171
- De Moor, Margriet 46, 95, 106, 107, 110, 112-115, 122, 171
*De schilder en het meisje (*The Painter and the Girl)* 46, 106, 107, 110, 112-114, 122, 171
- Deneire, Tom 75-76
- Den Hollander, Loes 43, 57, 59, 106
*Het scherventapijt (*A Carpet of Shards)* 106
- Deplouge, Jeroen 21
- De Sade, Marquis 69
- De Waal, Simon 46
*Een dief in de nacht (*A Thief in the Night)* 46
- Dewulf, Bernard 66, 133
*Kleine dagen (*Little Days)* 66, 133, 138, 168
- Dijkshoorn, Nico 54, 133, 138, 146, 147, 148, 153, 154
*Nooit ziek geweest (*Never Been Ill)* 133, 138, 146, 147, 148, 153, 154
- Doherty, David 112
- Donoghue, Emma 88, 95, 100, 101, 103-105, 122, 164, 165, 185
Kamer (Room) 88, 95, 100, 101, 103-105, 122, 164, 165, 185
- Doppen, Bas 34
- Dorrestein, Renate 35, 58, 95, 106, 107, 110, 112, 114, 115-121, 122, 162, 171, 173
*De blokkade (*Writer's Block)* 119
*Het geheim van de schrijver (*The Writer's Secret)* 119
*De leesclub (*The Reading Group)* 106, 110, 112, 114, 115-121, 122, 162, 171
- De stiefmoeder (*The Stepmother)* 106, 115-121, 122, 171, 173
- Onder de motorkap van het schrijverschap (*Under the Bonnet of Authorship)* 119
- De zondagmiddagauto (*The Sunday Afternoon Car)* 106, 107
- Dueñas, María 100
Het geluid van de nacht (The Time in Between) 100
- Durlacher, Jessica 106, 171
*De held (*The Hero)* 106, 171
- Dutton, Denis 17
- Eco, Umberto 45, 127
De begraafplaats van Praag (The Prague Cemetery) 45, 127
- Eder, Maciej 19, 20, 64, 101, 129, 172
- Eelbode, David 75-76
- Eggers, Dave 24, 128
The Circle 24
Wat is de wat (What is the What) 128
- El Bezaz, Naima 95, 106, 111
*Vinexvrouwen (*Middle-Class Women in Modern Suburbia)* 106, 111
- Elsschot, Willem 145
Kaas (Cheese) 145
- Empirical study of literature 27-28
- Enquist, Anna 66, 106, 171, 173
*De verdoovers (*The Anaesthetists)* 66, 106, 171, 173
- Enter, Stephan 45, 132, 171
*Grip (*Grip)* 45, 132, 171
- Evans, Nicholas 125, 129
De vergeving (The Brave) 125, 129
- Eve, Martin Paul 28
- Falcones, Ildelfonso 128, 129
De hand van Fatima (The Hand of Fatima) 128, 129
- Felski, Rita 39
- Fforde, Katie 46
Trouwplannen (Wedding Season) 46
- Fielding, Joy 101
Roerloos (Still Life) 101
- Field theory 10, 13-15
- Flanders 10, 38, 125, 189
- Foer, Jonathan Safran 97-98
Everything is Illuminated 97
Extremely Loud and Incredibly Close 97
- Follett, Ken 125, 129
Val der titanen (Fall of Giants) 125, 129
- Foregrounding 27
- Formalism 27
- Fortuin, Arjen 141, 143
- Fragoso, Margaux 100
Tijger, tijger (Tiger, Tiger) 100
- Franck, Julia 100
De middagvrouw (The Blind Side of the Heart) 100
- Franssen, Thomas 10, 80, 81

- Franzen, Jonathan 128, 165
De correcties (The Corrections) 128
Vrijheid (Freedom) 128, 165
- Galbraith, Robert 24
The Cuckoo's Calling 24
- Garrett, Sam 142, 146, 148
- Genootschap Onze Taal (the Dutch language society) 86, 131
- Gerrard, Nicci 100
Het weerzien (The Middle Place) 100
- Gilbert, Elizabeth 47, 95, 98-99, 101, 105, 122, 164, 165, 168, 184
Eten, bidden, beminnen (Eat, Pray, Love) 47, 95, 98-99, 101, 105, 122, 164, 165, 168, 184
Toewijding (Committed) 101, 105
- Giordano, Paolo 44, 128
De eenzaamheid van de priemgetallen (The Solitude of Prime Numbers) 44, 128
- Giphart, Ronald 133
*IJsland (*Iceland)* 133
*Zeven jaar goede seks (*Seven Years of Good Sex)* 133
- Grisham, John 23, 59,
- Groes, Sebastian 12, 197
- Grunberg, Arnon 66, 124, 132, 141, 148-156, 165, 192
Blauwe maandagen (Blue Mondays) 148
*Huid en haar (*Hair and Hide)* 66, 132, 148-156, 165
*De man zonder ziekte (*The Man Without Illness)* 66, 132, 148-156
*Selmonosky's droom (*Selmonosky's Dream)* 132
Tirza (Tirza) 148
- Haghighi, Anousheh 21
- Harbach, Chad 128
De kunst van het veldspel (The Art of Fielding) 128
- Hendriks, Pepijn 161
- Hennion, Antoine 15-16
- Herrmann, J. Berenike 25, 26-27
- Heydel, Magda 21
- Hildegard von Bingen 21
- Hill, Lawrence 128
Het negerboek (The Book of Negroes) 128
- Hodgkinson, Amanda 100
Britannia Road 22 (22 Britannia Road) 100
- Hoover, David L. 19, 21, 70, 78
- Hosseini, Khaled 44, 128
Duizend schitterende zonnen (A Thousand Splendid Suns) 44, 128
- Houellebecq, Michel 45, 58, 80, 127
De kaart en het gebied (The Map and the Territory) 45, 80, 127
- Intomart GfK 38
- Irving, John 59, 128
In een mens (In One Person) 128
De laatste nacht in Twisted River (Last Night in Twisted River) 128
- James, E.L. 23, 43, 46, 48, 63, 64-74, 161
Vijftig tinten grijs (Fifty Shades of Grey) 24, 46, 48, 63, 64, 66, 161
Vijftig tinten donkerder (Fifty Shades Darker) 43, 48, 67
Vijftig tinten vrij (Fifty Shades Freed) 43, 48, 67
Vijftig tinten-trilogie 13, 46, 48, 63, 66-77, 92, 101, 182, 192
- Janssen, Elisa 196
- Japin, Arthur 58, 59, 66, 132, 133, 171
*Dooi (*Thaw)* 66, 133
*De overgave (*Surrender)* 132, 171
*Vaslav (*Vaslav)* 132
- Jautze, Kim 159-160
- Jellema, Martine 98
fiskefet 142
- Jockers, Matthew L. 22-25, 70, 162
- Jonasson, Jonas 48, 79, 128
De 100-jarige man die uit het raam klom en verdween (The 100-Year-Old Man Who Climbed Out the Window and Disappeared) 48, 79, 128
- Journal of Computational Literary Studies (JCLS) 19
- Joyce, James 22
- Kestemont, Mike 20, 21
- Ketzan, Erik 25-27
- Keyword analysis 32, 55-61, 63, 85, 87, 95, 101, 121, 122, 124, 130, 140, 155, 174, 176, 190
- Kingsbury, Karen 101
Laatste dans (A Time to Dance) 101
Nooit te laat (Redemption) 101
- Kingsolver, Barbara 24
- Kinsella, Sophie 46, 47, 64, 146, 147, 148, 150, 153, 154
Ken je me nog? (Remember me?) 46
Mag ik je nummer even? (I've Got Your Number) 46, 47
Mini Shopaholic (Mini Shopaholic) 46
Shopaholic en baby (Shopaholic & Baby) 46, 146, 147, 148, 150, 153, 154
- Kluun, (Ray) 48, 124, 125, 133, 138, 141-148, 155
*Haantjes (*Alpha Males)* 48, 125, 133, 138, 141-148, 155
Komt een vrouw bij de dokter (Love Life) 141, 145
- Koch, Herman 44, 124, 133, 138, 141-148, 155
Het diner (The Dinner) 44, 48, 133, 142
Zomerhuis met zwembad (Summer House with Swimming Pool) 44, 48, 133, 138, 141-148, 155
- Koolen, Corina 12, 13, 33, 73, 92, 159, 164, 166, 177-181, 185, 197
- Koopman, Emy 178
- Krauss, Nicole 45, 96, 97, 100, 102, 107, 122, 127
Het grote huis (Great House) 45, 96, 97, 100, 102, 122, 127
- Kwantes, Johan 35
- Lanoye, Tom 45, 132, 134-135, 138, 155, 171, 172
*Heldere hemel (*Clear Sky)* 133, 138
Sprakeloos (Speechless) 45, 132, 134-135, 138, 155, 171, 172

- Lapidus, Jens 168
Snel geld (Easy Money) 168
- Larsson, Stieg 24, 44, 78
De Millennium-trilogie 24, 44, 48
Mannen die vrouwen haten (The Girl with the Dragon Tattoo) 44, 48
De vrouw die met vuur speelde (The Girl Who Played with Fire) 44
- Launspach, Rik 125, 133
*1953 (*1953)* 125, 133
- Lauwers, Jeroen 75-76
- Le Carré, John 125
Ons soort verrader (Our Kind of Traitor) 125
- Lensink, Saskia 84
- Letterenfonds (Dutch Foundation for Literature) 107, 111, 116, 141, 148
- Lewinsky, Charles 128
Het lot van de familie Meijer (Melnitz) 128
*De verborgen geschiedenis van Courtillon (*Saint John's Eve)* 128
- litRiddle R Package 12, 33, 52, 55, 66, 67, 68, 82, 96, 97, 98, 126, 143, 170, 190, 204, 214
- LIWC (Linguistic Inquiry and Word Count) 180
- Lu, Kristine 21
- Luijten, Clementine 77, 78
- MacDowell, Heather & Rose 96
Diner voor 2 (Turning Tables) 96
- Mak, Geert 37, 58
Reizen zonder John (In America: Travels with John Steinbeck) 37
- Mankell, Henning 59, 63, 77-80
De Chinees (The Man from Beijing) 77-80
*De Daisy Sisters (*Daisy Sisters)* 77-80
De gekweldde man (The Troubled Man) 77-80
De geschiedenis van een gevallen engel (A Treacherous Paradise) 77-80
Kennedy's brein (Kennedy's Brain) 77-80
- Mansell, Jill 46, 64
Drie is te veel (Two's Company) 46, 64
Versier me dan (Take a Chance on Me) 46, 64
- Marlantes, Karl 128
Matterhorn (Matterhorn) 128
- Mastras, George 129
Tranen over Kashmir (Fidali's Way) 129
- Matthew effect 90
- McCarthy, Rachel 21, 114
- McCoy, Sarah 100
De bakkersdochter (The Baker's Daughter) 100
- Meyer, Deon 43
- Middelbeek-Oortgiesen, Janny 77
- Miedema, Niek 131-132
- Mitchell, David 128, 131, 155
De niet verhoorde gebeden van Jacob de Zoet (The Thousand Autumns of Jacob de Zoet) 128, 131, 155
- Mixed Grill* 34
- Moens, Sara 21
- Monty Python's Flying Circus 142
- Morrison, Toni 24
- Mortier, Erwin 45, 132, 134-135, 138, 146, 147, 153, 154, 155, 171, 172
Gestameld liedboek (Stammered Songbook) 45, 132, 134-135, 138, 155, 171
Godenslaap (While the Gods Were Sleeping) 45, 132, 134, 138, 146, 147, 153, 154, 171
- Mulisch, Harry 54, 56
- Murakami, Haruki 45, 59, 64, 65, 127, 128, 129, 165, 193
1984 (1Q84) 128, 165
Norwegian Wood (Norwegian Wood) 45, 64, 65, 127, 129, 165
- Nagelhout, Erica 33
- Nash Equilibrium 75-76
- Nash, John Forbes 75
- NBD Biblion 34, 42
- Neill, Fiona 168
Vriendschap, liefde en andere stomiteiten (Friends, Lovers and Other Indiscretions) 168
- Neo-Darwinism 9, 13, 17-18
- Netherlands 9-11, 28, 33, 34, 37, 38, 39, 42, 54, 60, 63, 80, 83, 90, 95, 116, 125, 137, 138, 146, 148, 149, 162, 177, 188, 189, 193, 194, 195, 197-198
- New York Times, The* 23-24
- Nieuwland, Ilja 35
- Noort, Saskia 37, 44, 48, 72
De eetclub (The Dinner Club) 37, 44
*Koorts (*Fever)* 48
- NRC-next 143, 148
- NUR (Dutch Uniform Classification) 36, 42, 43, 45
- Organisation for Economic Co-operation and Development (OECD) 196
- O'Sullivan, James 21, 114
- PCA (principal components analysis) 63, 69-70, 71, 72, 78, 79, 86, 87, 88, 95, 101, 102, 105, 110, 114, 116, 124, 129, 130, 139, 155, 172, 173, 191
- Pearl, Lisa 21
- Peer, Willie van 27-28
- Peetz, Monica 101
*De dinsdagvrouwen (*Tuesday Women)* 101
- Pennebaker, James 180
- Peters, Arjan 142, 143
- Pick, Alison 100
Donderdagskind (Far to Go) 100
- Picoult, Jodi 100
Negentien minuten (Nineteen Minutes) 100
- Piper, Andrew 28
- Post van der Linde, Carole 77
- Programme for International Student Assessment (PISA) 196, 197
- Radway, Janice A. 73, 74
- Ravelli 133
*De Vliegenvanger (*Fly Trap)* 133

- Richardson, Samuel 21
Clarissa 21
- Riddell, Allen 39-40
- Rosenboom, Thomas 132, 171
Mechanica (**Mechanics*) 132
Zoete mond (**A Mouth Full of Sweets*) 132, 171
- Rosnay, Tatiana de 44, 48, 51, 52, 81, 100
Het appartement (**The Neighbour*) 100
Het huis waar jij van hield (*The House I Loved*) 100
Kwetsbaar (*Moka*) 100
Haar naam was Sarah (*Sarah's Key*) 44, 48, 51, 52, 100
Die laatste zomer (*A Secret Kept*) 81, 100
- Roth, (Joseph?) 54
- Rowling, J.K. 24, 52, 80
Harry Potter en de Relieken van de Dood (*Harry Potter and the Deathly Hallows*) 80
- Ruiz Zafón, Carlos 44, 80, 81, 128
De gevangene van de hemel (*The Prisoner of Heaven*) 128
De schaduw van de wind (*The Shadow of the Wind*) 44, 81, 128
Het spel van de engel (*The Angel's Game*) 80, 128
- Rybicki, Jan 20, 21, 114
- Sansom, C.J. 125
Winter in Madrid (*Winter in Madrid*) 125
- Schiffers, Kyra 34
- Schöch, Christof 25-26
- Scholten, Jaap 133
Kameraad Baron (*Comrade Baron*) 133
- Schwarz, Ori 16-17
- Sebald, W.G. 54
- Shah, Hannah 100
De dochter van de imam (*The Imam's Daughter*) 100
- Siebelink, Jan 56, 133
Het lichaam van Clara (**Clara's Body*) 133
Oscar (**Oscar*) 133
- Siemens, Renske 34
- Slee, Carry 166
Fatale liefde (**Fatal Love*) 166
- Slob, Marjan 73-75
- Smeets, Mart 66, 124, 125, 133, 136-138, 155, 162, 165, 184, 194
De afrekening (**The Reckoning*) 66, 125, 133, 136-138, 155, 162, 165, 184
Een koud kunstje (**A Breeze*) 66
- Smit, Susan 106, 112, 114, 171
Vloed (**High Tide*) 106, 112, 114, 171
- Smits, Manon 103
- SpaCy 168
- Spijker, Rita 106
Tussen zussen (**Between Sisters*) 106
- Spinoza 57, 60
- Springer, F. 56, 57, 133, 193
Quadrige (**Quadrige*) 133
- Squires, Claire 162, 182, 183
- Steel, Danielle 23, 24
- Stheeman, Tjadine 96
- Stockett, Kathryn 81, 100, 103, 165
Een keukenmeidenroman (*The Help*) 81, 100, 103, 165
- Structuralism 27
- Stylo Package for R 20, 64, 70, 86, 101, 114, 129, 137, 147, 172, 190
- Stylometry 18, 19
- Surdèl, Steven 35
- Sybesma, Edith 77
- 't Hart, Maarten 133
Verlovingstijd (**Betrothal Time*) 133
- Thomése, P.F. 66, 132, 138
De weldoener (**The Benefactor*) 66, 132, 138
- Topic modelling 22-23, 158, 162-163, 165-167, 179, 184, 194
- Treur, Franca 106, 171, 173
Dorsvloer vol confetti (**Confetti on the Threshing Floor*) 106, 171, 173
- Twain, Mark 22
- Underwood, Ted 28
- United Kingdom 10, 12, 37, 43, 81, 126, 197
- United States 10, 23, 81, 90, 197
- Van Beijnum, Kees 133, 138
Een soort familie (**A Kind of Family*) 133, 138
- Van Bree, Corry 77, 79
- Van Cranenburgh, Andreas van 12, 13, 33, 92, 158-167, 170, 180, 184, 185, 190, 191
- Van der Heijden, A.F.Th. 45, 48, 132, 171
Tonio (**Tonio*) 45, 48, 132, 171
- Van der Kwast, Ernest 133
Mama Tandoori (*Mama Tandoori*) 133
- Van der Meer, Vonne 106, 171
Het zingen, het water, de peen (**Carolling, Water, Carrot*) 106
De vrouw met de sleutel (**The Woman with the Key*) 106, 171
- Van der Veer, Rob 96
- Van der Wouden, Ton 184
- Van Dijk, Yra 152-153, 156, 197
- Van Dis, Adriaan 132
Tikkop (*Betrayal*) 132
- Van Galen, Alex 113, 138
Süskind (**Süskind*) 113, 138
- Van Gastel, Chantal 153, 154
Geknipt voor jou! (**Perfect for You!*) 153, 154
- Van Gelder, Mariëtte 150
- Van het Reve, Karel 11, 14, 55, 191-193
- Van Kooten, Kees 58
- Van Meegeren, Han 17
- Van Rossum, Lisanne 85, 176
- Van Royen, Heleen 106, 108-109, 122, 136
De manmentester (**The Tester of Men*) 106, 108-109, 122, 136
Sabine (**Sabine*) 106

- Van Sambeek, 106, 162
Koninginnenrit (*Queen's Ride) 106, 162
- Van Sambeek, Liza 106
Bloed, zaad en tranen (*Blood, Seed, and Tears) 106
- Van Wingerden, Wouter 161
- Van Zundert, Joris 21, 168
- Verhoef, Esther 106
Tegenlicht (*Backlight) 106
- Verhulst, Dimitri 133
De laatste liefde van mijn moeder (*The Last Love of My Mother) 133
- Vermeer, Johannes 17
- Vermeer, Suzanne 43, 66
Après-ski (*Après Ski) 66
Cruise (*Cruise) 66
- Verstraeten, Pieter 137
- Vincent, Paul 134, 135
- Vlek, Ronald 126
- Vlugt, Simone van der 106, 112
Jacoba, dochter van Holland (*Jacoba, Daughter of Holland) 106, 112
Rode sneeuw in december (*Red Snow in December) 106, 112
- Volkskrant, de* 142, 148
- Von Heydebrand, Renate 40
- Voskuijl, Anouschka 96, 165
Dorp (*Village) 96, 165
- Voskuil, J.J. 56, 66, 133, 138-140, 161, 165
Het Bureau 140
De buurman (*The Neighbour) 66, 133, 138-140, 161, 165
- Vuijsje, Robert 125, 133, 138
Alleen maar nette mensen (*Decent People Only) 125, 133, 138
- Weijts, Christiaan 35
- Weiner, Jennifer 66
Sommige meisjes (*Certain Girls) 66
- Weir, Andy 25-26
The Martian 25-26
- Weisberger, Lauren 46, 64
Champagne in Chateau Marmont (*Last Night at Chateau Marmont) 46, 64
Chanel Chic (*Chasing Harry Winston) 46, 64
- Wieringa, Tommy 132, 166, 171
Caesarion (*Caesarion) 132, 166, 171
- Winko, Simone 40
- Winter, Leon de 133
Recht op terugkeer (*The Right to Return) 133
VSV of daden van onbaatzuchtigheid (*Acts of Selflessness) 133
- Wolff-Bekker, Elizabeth 21
- Worthy, James 56, 125, 133
James Worthy (*James Worthy) 125, 133
- Whiteside, Shaun 141
- Woolf, Virginia 21
Night and Day 21
- Yalom, Irving D. 60, 128
Het raadsel Spinoza (*The Spinoza Problem) 60, 128
- Zwaan, Josha 106, 112, 171
Parnassia (*Parnassia) 106, 112, 171
- Zwagerman, Joost 133
Duel (*Duel) 133

W

hat is literature? Can we measure 'literariness' in texts themselves? The innovative Computational Humanities project 'The Riddle of Literary Quality' asked thousands of Dutch readers for their opinion about contemporary Dutch and translated novels. The public shared which novels they had read, what they really thought of them, and how they judged their quality. Their judgments of the same novels were compared with the results of computational analysis of the books.

Using evidence from almost 14,000 readers and building on more textual data than ever before, Van Dalen-Oskam and her team uncovered unconscious biases that shed new light on prejudices many people assumed no longer existed. This monograph explains in an accessible way how the project unfolded, which methods were used, and how the results may change the future of Literary Studies.

Karina van Dalen-Oskam is head of the Computational Literary Studies research group at Huygens Institute for the History and Culture of the Netherlands, an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW). She is also professor in Computational Literary Studies at the University of Amsterdam, and Honorary Professor in the Humanities at the University of Wolverhampton (UK).

AUP.nl