



K O N I N K L I J K E N E D E R L A N D S E
A K A D E M I E V A N W E T E N S C H A P P E N

Personal names in literature: a quantitative approach.

van Dalen-Oskam, K.H.

published in

Acta Academiae Regiae Scientiarum Upsaliensis
2013

document version

Peer reviewed version

document license

CC BY

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

van Dalen-Oskam, K. H. (2013). Personal names in literature: a quantitative approach. *Acta Academiae Regiae Scientiarum Upsaliensis*, 22, 153-168.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

pure@knaw.nl

Karina van Dalen-Oskam

Huygens Institute for the History of the Netherlands (Royal Netherlands Academy of Arts and Sciences), The Hague

Personal names in literature: a quantitative approach¹

1 Introduction

Is it possible to write a novel that does not have a single personal name in it? Or better perhaps: a *readable* novel? Undoubtedly there are more, but I know of only one that does not contain any names, be it personal, geographical, or other: *Blindness*, a novel by the Portuguese 1998 Nobel Prize winner José Saramago. The novel describes how an unexplained mass epidemic of blindness afflicts nearly everyone in an unnamed city and depicts the horrible social breakdown that follows. The epidemic ends just as sudden and unexplained as it started. The characters in this novel do not get a name, but they are described with a noun phrase referring to what they are: the doctor, the doctor's wife, the first blind man, the old man with the black eyepatch, the girl with dark glasses, the boy with the squint.² Formally these noun phrases are not names but they are uniquely referring in the story world of the novel and thus share an important function with names. The avoidance of names has a clear function in the story, as can be illustrated with a quotation from the first part of the novel. Most of the main characters, being the very first group of people that went blind, have been placed in quarantine in an asylum. When a new group of blind men and women arrive in their ward,

The doctor's wife remarked, It would be best if they could be counted and each person gave their name. Motionless, the blind internees hesitated, but someone had to make a start, two of the men spoke at once, it always happens, both then fell silent, and it was the third man who began, Number one, he paused, it seemed he was about to give his name, but what he said was, I'm a policeman, and the doctor's wife thought to herself, He didn't give his name, he too knows that names are of no importance here. Another man was introducing himself, Number two, and he followed the example of the first man, I'm a taxi-driver. (Saramago 2005:57)

Many readers find this novel unreadable, but not because of the lack of names. As can be gathered from the above quotation, Saramago has a very personal style with long sentences, commas in stead of periods, and a very loose demarcation of dialogue, which does not sit well with many readers but is much appreciated by the literary hedonists.

This leads us to the question this contribution will deal with: What part do names play in novels in general, especially personal names? How many times do they occur and what are their functions? I will approach this topic from a quantitative point of view and zoom in on some interesting novels in my pilot corpus.³

2 Names in novels

For my research I gathered a digital corpus of 22 Dutch novels and 22 English novels. Four of each of the 22 novels were written for children/young adults. The corpus included 10 translations into the other language in both sets. The Dutch corpus has 20 novels from the Netherlands and two from Flanders. The English corpus contains 13 novels from the United States, 5 from Great-Britain and 4 from Canada. The focus of the corpus is on the last 20 years. Since a lot of the work still has to be done manually,⁴ the corpus is admittedly rather small (44 plus 20 novels), but it does present some rather interesting first observations.

| Sigla | Author <i>Translator</i> | Title (year) |
|------------------|------------------------------------|---------------------------------------|
| Abdolah_OR | Kader Abdolah | Spijkerschrift (2000) |
| Abdolah_TR | <i>Susan Massotty</i> | My father's notebook (2006) |
| Amis_OR | Martin Amis | Time's arrow (1991) |
| Amis_TR | <i>Gideon den Tex</i> | De pijl van de tijd (1991) |
| Atwood_OR | Margaret Atwood | The handmaid's tale (1985) |
| Atwood_TR | <i>Gerrit de Blaauw</i> | Het verhaal van de dienstmaagd (1987) |
| Bakker_OR | Gerbrand Bakker | Boven is het stil (2006) |
| Bakker_TR | <i>David Colmer</i> | The twin (2009) |
| Barrett | Andrea Barrett | The forms of water (1993) |
| Bernlef | J. Bernlef | Hersenschimmen (1984) |
| Boon | Louis Paul Boon | Menuet (1948) |
| Byars | Betsy Byars | Bingo Brown, Gypsy Lover (1990) |
| Chabon | Michael Chabon | The final solution (2004) |
| Daem | Geertrui Daem | Koud (2001) |
| Doctorow_OR | E. L. Doctorow | The march (2004) |
| Doctorow_TR | <i>Sjaak Commandeur</i> | De mars (2006) |
| Dorrestein_OR | Renate Dorrestein | Verborgen gebreken (1996) |
| Dorrestein_TR | <i>Hester Velmans</i> | A crying shame (2003) |
| Dragt | Tonke Dragt | De brief voor de koning (1962) |
| Enright_OR | Anne Enright | The gathering (2007) |
| Enright_TR | <i>Piet Verhagen</i> | De samenkomst (2007) |
| Erdrich | Louise Erdrich | The plague of doves (2008) |
| Ferron | Louis Ferron | Het stierenoffer (1975) |
| Foer_OR | Jonathan Safran Foer | Everything is illuminated (2002) |
| Foer_TR | <i>Peter Abelsen</i> | Alles is verlicht (2002) |
| Glastra van Loon | Karel Glastra van Loon | Lisa's adem (2001) |
| Golding | William Golding | The inheritors (1955) |
| Haasse | Hella S. Haasse | Fenrir (2000) |
| Hermans_OR | Willem Frederik Hermans | Nooit meer slapen (1966/1993) |
| Hermans_TR | <i>Ina Rilke</i> | Beyond sleep (2006) |
| Japin_OR | Arthur Japin | De zwarte met het witte hart (1997) |
| Japin_TR | <i>Ina Rilke</i> | The two hearts of Kwasi Boachi (2000) |
| LeGuin | Ursula LeGuin | Jane on her own (2003) |

| | | |
|--------------|-----------------------------|-----------------------------------|
| L'Engle | Madeleine L'Engle | A wrinkle in time (1962) |
| Loon | Paul van Loon | Dolfje weerwolfje (1996) |
| MacLeod | Alistair MacLeod | No great mischief (1999) |
| McCarthy_OR | Cormac McCarthy | The road (2006) |
| McCarthy_TR | <i>Guido Goluke</i> | De weg (2010) |
| McEwan_OR | Ian McEwan | On Chesil Beach (2007) |
| McEwan_TR | <i>Rien Verhoef</i> | Aan Chesil Beach (2007) |
| Moring_OR | Marcel Möring | In Babylon (1997) |
| Moring_TR | <i>Stacey Knecht</i> | In Babylon (1999) |
| Morrison_OR | Toni Morrison | The bluest eye (1970) |
| Morrison_TR | <i>Nettie Vink</i> | Het blauwste oog (1984) |
| Mulisch_OR | Harry Mulisch | De aanslag (1983) |
| Mulisch_TR | <i>Claire Nicolas White</i> | The assault (1985) |
| Nooteboom_OR | Cees Nooteboom | Het volgende verhaal (1991) |
| Nooteboom_TR | <i>Ina Rilke</i> | The following story (1994) |
| Ondaatje_OR | Michael Ondaatje | In the skin of a lion (1987) |
| Ondaatje_TR | <i>Graa Boomsma</i> | In de huid van een leeuw (1989) |
| Peper | Rascha Peper | Vingers van marsepein (2008) |
| Proulx | E. Annie Proulx | Postcards (1992) |
| Rice | Anne Rice | Interview with the vampire (1976) |
| Roth | Philip Roth | The human stain (2000) |
| Terlouw | Jan Terlouw | Oorlogswinter (1972) |
| Thomese_OR | P. F. Thomése | Schaduwkind (2003) |
| Thomese_TR | <i>Sam Garrett</i> | Shadow child (2005) |
| Updike_OR | John Updike | Of the farm (1966) |
| Updike_TR | <i>Frans Bijlsma</i> | Terug en verder (1967) |
| Vlugt | Simone van der Vlugt | Schuld (2007) |
| Walsh | Jill Paton Walsh | A chance child (1978) |
| Wieringa | Tommy Wieringa | Alles over Tristan (2002) |
| Wolkers_OR | Jan Wolkers | Een roos van vlees (1963) |
| Wolkers_TR | <i>John Scott</i> | A rose of flesh (1967) |

Table 1: The analyzed novels. The extension _OR stands for 'Original', _TR stands for 'Translation'

There are several ways to look at the usage and functions of names in literary texts. The first thing I wanted to know was how many names usually occur in a novel, being personal names, geographical names, and other names such as buildings, events, book titles, animal names. I chose to only consider the 'prototypical' names, meaning those names that are considered to be names by readers in general. The main criterium is that a name refers to a unique person, place, or object. This meant the exclusion of currencies, days of the week, months, etc. The bar chart in Fig. 1 shows the total percentage of tokens that are a name or part of a name in the forty-four original novels (I will not go into the 20 translations in this article):

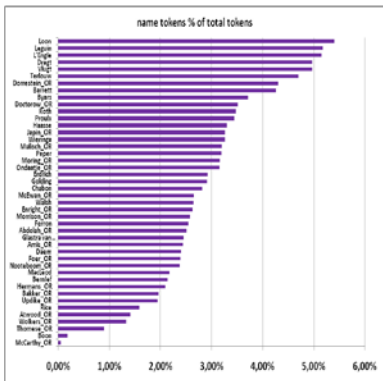


Fig. 1: % all name tokens

[file: Personalnames_Karina_fig_1.png (colour)]

[file: Personalnames_Karina_fig_1_BLWH.png (black-and-white)]

The bar chart in Fig. 1 shows that there is a small group having more than 3.5 % of name tokens and a small group having less than 2 %. The novels written in the two different languages nicely mix in this list. The remainder of this contribution will go into the novels with the highest percentages of name tokens, but first we will go deeper into the novel which has the lowest name token percentage: Cormac McCarthy's *The Road*.

3 Cormac McCarthy, *The Road*

The novel is about a father and his young son who travel south on the road, hoping to flee from the ever worse winters after a huge catastrophe destroyed most of human life and all of the flora and fauna. In flashbacks we learn that the son was born a couple of days after the catastrophe and that somewhere in time his mother could not handle the rough life anymore and went away. In the story world this is suggested to mean that she committed suicide, as so many other people are shown to have done. The only food there is, are canned goods in old stores or in hidden bunkers. People who survived shun other people, or they gang up and kill others to eat them. The father tries his best to set a good example for his son even during these impossible times but eventually fails.

There are two personal names in this novel (in the quotations I will systematically have the names printed in bold font). The first is 'God', when referring to the godly being. When the word *God* is used as a curse, I have not included it in my countings. God is used as a personal name in for instance the next couple of quotations: 'He [the father] knew only that the child was his warrant. He said: If he is not the word of **God God** never spoke.' (McCarthy 2006:5). To his son he says: 'You wanted to know what the bad guys looked like. Now you know. It may happen again. My job is to take care of you. I was appointed to do that by **God**. I will kill anyone who touches you. Do you understand?' (McCarthy 2006:77)

In the next quote, 'God' can be either interpreted as a name or as an invocation, a curse. It is representative of the bad light in which God is seen throughout this novel:

Then he [the father] just knelt in the ashes. He raised his face to the paling day. Are you there? he whispered. Will I see you at the last? Have you a neck by

which to throttle you? Have you a heart? Damn you eternally have you a soul? Oh **God**, he whispered. Oh **God**. (McCarthy 2006:11-12)

The novel has one other personal name in it. The father and son meet an old man, and, spurred on by the boy, decide to try to help him and give him something to eat. (The dialogue in this novel is presented without any quotes or explicit turntakings, which means that the reader has to actively keep track of who is speaking.)

What's your name?

Ely.

Ely what?

What's wrong with **Ely**?

Nothing. Let's go. (McCarthy 2006:167)

Since this is the only character in the novel that has a name, we expect it to have a special significance. The first association that comes to mind is a Biblical one: Ely is the name of an Old Testament priest who is especially known for raising the young prophet Samuel. Ely had two sons who committed many crimes. When the ark of God was stolen, his sons were slain. Ely dies when he hears the news - he is said to be 98 years old and blind by then (1 Samuel 4, 15). The Ely in *The Road* is also described as being (very) old (he states he is ninety) and as almost blind. When the dialogue continues, the correctness of the Biblical association seems to be confirmed:

How would you know if you were the last man on earth? He [the father] said.

I dont guess you would know it. You'd just be it.

Nobody would know it.

It wouldnt make any difference. When you die it's the same as if everybody else did too.

I guess **God** would know it. Is that it?

There is no God.

No?

There is no God and we are his prophets. (McCarthy 2006:169-170)

After this, we find a so-called "meta-onomastic" statement of the old man about the function of names, prompted by a question the father asks him:

Is your name really **Ely**?

No.

You dont want to say your name.

I dont want to say it.

Why?

I couldnt trust you with it. To do something with it. I dont want anybody talking about me. To say where I was or what I said when I was there. I mean, you could talk about me maybe. But nobody could say that it was me. I could be anybody. I think in times like these the less said the better. If something had happened and

we were survivors and we met on the road then we'd have something to talk about. But we're not. So we dont. (McCarthy 2006:171-172)

'The less said the better': in a post-apocalyptic world, names are irrelevant and only remind one of things lost forever. Names lead to talk, and talk is useless when things have gone catastrophically wrong.

The function of (the lack of) names in *The Road* thus is comparable to the lack of names in *Blindness*, as described in Section 1 of this contribution.

4 Zooming in on personal names

After this sinister excursion, in which the lack of names proved to have a clear function in the novel by Cormac McCarthy, we go back to the bar chart in Fig. 1 with the overview of name token percentages. A closer look at this chart reveals that the six novels with the highest percentages are all children's books and books for young adults (Loon, LeGuin, L'Engle, Dragt, Vlugt, Terlouw). The seventh (Byars) of the eight children's book is not far lower in the ranking, at rank 9, and the eight (Walsh) is rather alone a bit below the middle of the whole list of novels, at rank 24. My hypothesis is that children's books have more names in them because the authors try to avoid personal pronouns and instead use personal names to keep the story easier to understand.

If we turn this around, the question now arises what the function is of the higher frequencies of name tokens in the two novels for *adults* that the bar chart shows *between* the sixth and seventh of the children's books, and who both (Dorrestein and Barrett) have more than 4 % name tokens. But before we have a closer look at these novels, I want to present some further visualizations of the name use in my corpus of 44 novels. Since in this volume of articles personal names are the main focus, in Fig. 2 I made the three main name types distinguishable in the bars as well: personal names (purple), geographical names (red) and other names (green). The list is sorted from the highest percentage of personal name tokens at the top to the lowest at the bottom of the chart, and when we take into account the frayed edge to the right, it will be clear that several novels changed places in the chart compared to Fig. 1 in which the complete name token percentages were ranked.

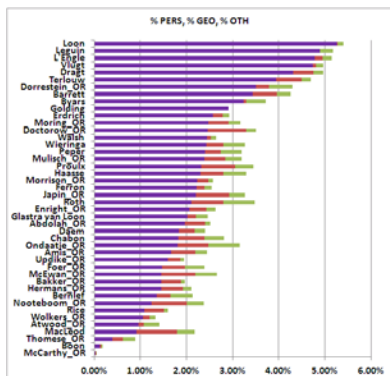


Fig. 2: all name tokens, three main categories

[file: Personalnames_Karina_fig_2.png (colour)]

[file: Personalnames_Karina_fig_2_BLWH.png (black-and-white)]

If we again try to locate the eight books for children and young adults, we find that the same six still are at the top of the ranking, although with minor place changes. The seventh (Byars) again comes after the same two novels for adults as in Fig. 1. The eighth (Walsh) has risen in the ranks, and instead of at rank 24 is now at rank 14.

The next distinction to be analyzed concerns the one between names referring to fictional and names referring to real persons. Personal names in a novel can refer either to fictional characters that only occur in the story world of the novel, or to persons existing or having existed in the real world. I call the first group *plot internal* names and the second *plot external* names for short – the correct turn of phrase would be ‘names referring to plot internal/external entities’. Examples of plot internal names are personal names such as **Frodo** and **Sam** in *The lord of the rings* by J. R. R. Tolkien. An example of a plot external personal name would be if in a novel we read that one of the characters highly values the poetry of **Tomas Tranströmer**. This is a handy way for the author to characterize the fictional person in his/her likes and dislikes.

If we distinguish all personal names in the 44 novels into plot internal and plot external names, we find that the plot internal names almost always occur much much more than plot external names. The ranking in Fig. 3 is now based on the percentage of the plot internal name tokens.

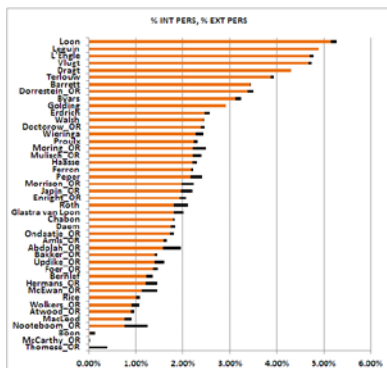


Fig. 3: % all personal name tokens, plot internal (orange) and plot external (black)]

[file: Personalnames_Karina_fig_3.png (colour)]

[file: Personalnames_Karina_fig_3_BLWH.png (black-and-white)]

If we mark the 8 books for children / young adults again, we find the same picture as before: the same six at the top, followed by the seventh (Byars) after the same two novels for adults (but now in the order of Barrett and Dorrestein), and the eighth one (Walsh) moving up again in the ranks, now from rank 14 to rank 12.

5 Andrea Barrett, *The Forms of Water*, and Renate Dorrestein, *A Crying Shame*

Let's have a closer look now at the two books for adults that have a frequency of (plot internal personal) name tokens which is comparable to the books for children and young adults in the pilot corpus. It is to be expected that the higher frequencies of names in these two novels have another function than the one assumed for the books for children and young adults. What is that function or what are these functions? Table 2 presents more details about the top-12 novels in the last bar charts.

| Sigla | Author | Title (year of publication) |
|-------------------|----------------------|------------------------------------|
| Loon | Paul van Loon | Dolfje weerwolfje (1996) |
| LeGuin | Ursula LeGuin | Jane on her own (2003) |
| L'Engle | Madeleine L'Engle | A wrinkle in time (1962) |
| Vlugt | Simone van der Vlugt | Schuld (2007) |
| Dragt | Tonke Dragt | De brief voor de koning (1962) |
| Terlouw | Jan Terlouw | Oorlogswinter (1972) |
| Barrett | Andrea Barrett | The forms of water (1993) |
| Dorrestein | Renate Dorrestein | Verborgten gebreken (1996) |
| Byars | Betsy Byars | Bingo Brown, Gypsy Lover (1990) |
| Golding | William Golding | The inheritors (1955) |
| Erdrich | Louise Erdrich | The plague of doves (2008) |
| Walsh | Jill Paton Walsh | A chance child (1978) |

Table 2: The top-12 novels according to the frequency of plot internal personal name tokens

The two most significant 'interlopers' are Andrea Barrett's *The forms of water* and Renate Dorrestein's *Verborgten gebreken*, translated into English as *A crying shame*. Both novels have a story line that focuses on family relations. *The forms of water* is about the Auberon family. Main character is Henry Auberon, a failed property developer, divorced from his wife and hardly ever seeing his two children. Henry and his sister Wiloma grew up with their grandparents after their parents died in a car crash. Their uncle Brendan was a monk then at the other side of the world. But Brendan lost his faith and now, old and wheelchair-bound, lives close to his family again. He gets Henry to drive him to Paradise Valley, the land where they both grew up and which was long ago changed into a water reservoir. Wiloma and her ex-husband go after them, just like the almost adult children of both Henry and Wiloma. They all meet at the reservoir.

The four teenagers travel together and we see how complex their relations are (it is no different with the adults, by the way).

Delia smiled at **Roy** and teased him quietly. But there was no mistaking the confident way she touched his arm.
Mine, that touch said, and **Wendy** convinced herself that **Delia** deserved him.
Roy gazed over **Delia's** head at **Wendy** and stroked **Delia's** hair quite deliberately. **Wendy** felt that he did this not to be mean but to let her know where they stood. (Barrett 1993:226)

At the end of the novel the whole family converge at the waterfront where Brendan has himself rowed to the spot of the church of his youth.

Brendan had kept his back toward **Wiloma** and the sounds **Henry** was making in the distance, but when he heard his nickname he craned his head over his shoulder and saw **Lise** there, and **Delia** and **Wendy** and **Win**, lined up on the shore and waving at him. For a minute he thought he might be dreaming. He hadn't seen them all together in years; **Wendy** and **Win** took turns coming to visit him with their mother, and although he'd seen **Lise** only yesterday, before that he hadn't seen her or **Delia** since **Delia** had left for college. (Barrett 1993:258)

One could perhaps state that the high frequency of name use logically follows from the author throwing so many people together at the same place and having them travel together to the same destination. But often the repetition of a name does feel a bit forced. This leads me to the hypothesis that the more frequent explicitation with a name instead of a pronoun in this novel is a stylistic procedure or 'trick' if you will to highlight the oppressive complexity of this family's life and history together (and apart). The novel does not contain any revealing meta-onomastic statements that can be related to the high frequency of the name tokens, however.

A crying shame (I will quote from the 2003 translation of Dorrestein 1996) is about 70-year old Agnes, a single woman who had four brothers, one of which she clearly was in love with, and about a 10-year old girl named Christine, who wants to be called Chris. Chris is on a holiday in Scotland with her mother, stepfather, and her older and younger brother. The older one sexually abuses her. Chris and her younger brother somehow run away and end up on the old holiday farm of Agnes's family, where Agnes just visits for one last time, looking back at her long life in a family with many nephews and nieces and with brothers which each had their specific problems.

Now that **Justus**, **Frank**, **Robert** and **Benjamin** are no longer there to break people's legs or to smack the sly grins off their faces, now they're all laughing at her. And who knows what else they may be up to, behind her back. "Say, **Joyce**," **Gemma** may have remarked with motherly foresight, "what if you staked a claim to that flat of Aunt **Agnes's**? (...)"

Joyce, eager but somewhat guilty, to her great-aunt **Elise**: "Couldn't she just stay in Mull?"

Elise, alarmed, to her nephew **Joost**, a solicitor: "I've got this feeling **Agnes** might be intending to stay in Port na Bà. Can't I, how do you call it, evict her?" (Dorrestein 2003:119-120)

Chris's family is rather small, and it does not seem to be the case that in the episodes concerned with Chris and her little brother Tommy there is a more frequent use of names. It's Agnes's extended family, and the way all their names are listed in Agnes's thoughts, that seem to be responsible for the higher frequency of name tokens in this novel. Agnes reminisces about her own role in her large family, repeatedly listing persons by their names and roles (brothers, sisters-in-law, nephews and nieces she took care of and who now do not pay any attention to her). So again it is the oppressiveness of family that

seems to be expressed by the repetition and high-frequency of the names of all those family members, especially of Agnes’s brothers:

There’s **Justus**: thoughtful, an irritatingly slow worker, but reliable, and so even-tempered. **Frank**: sketching on the back of an envelope, making plans – if we were to knock down the chimney, no, I mean it, just hear me out, and then we did something about the f-façade – is anyone listening to me? **Robert**: no-one’s having a drink until we’re finished with that last wall, that last skirting board, not until then. You’ll all thank me later. And **Benjamin**: come on **Agnes**, everyone has to spend at least twenty years scraping paint off the woodwork, you’re getting off easy here. (Dorrestein 2003:71)

Other methods will have to be developed to find out whether this hypothesis is correct. One option could be to mark a digital text into episodes primarily focusing on Agnes and primarily focusing on Chris and Tommy, and calculate the name token percentages in each of these episodes. Another possibility that could be explored is to find out the level of nearness of names in the two different kinds of episodes; we would then be able to see whether the ‘name lists’ are indeed mostly present in the episodes relating to Agnes.

6 Categories of personal names and amount of different characters

In the quotations in Section 5 all personal names were first names. To find out whether first names are generally more frequent than other categories of personal names in the corpus, we have a closer look at the amount of different categories of personal names. I have distinguished between first names, family names, and nicknames (of which it is often unclear whether they function as a replacement of a first name or a family name). Fig. 4 shows how these name categories are distributed in the analyzed novels, sorted from the highest amount of name tokens for first names to the lowest.

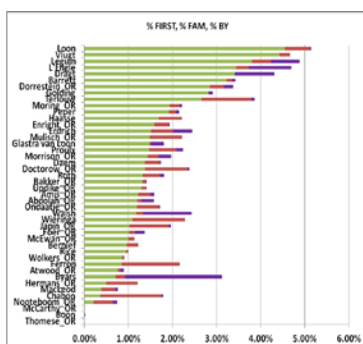


Fig. 4: all personal name tokens, three main categories
[file: Personalnames_Karina_fig_4.png (colour)]
[file: Personalnames_Karina_fig_4_BLWH.png (black-and-white)]

The group of books for children and young adults starts to break up slightly. The first five at the top are from this group. The other three are at rank 9, 27, and 37. The bar chart shows that the ones at rank 27 and 37 have a large amount of name tokens being a

nickname. Both of these novels contain main characters who are primarily referred to by their nickname.

Apart from the absolute numbers of name tokens and name occurrences, it could be useful to also look at the amount of different characters in a novel that are referred to with a name. In this approach we overlook the number of name tokens and the number of different names referring to the different characters. Fig. 5 shows the number of different plot internal, so ‘acting’ characters’, names in the corpus of 22 Dutch novels (the English corpus does not have that information yet). What we then see, if we sort the novels from the one with the highest amount of named entities (characters) to the lowest, is the following – and finally we do *not* see the children’s books at the top of the list:

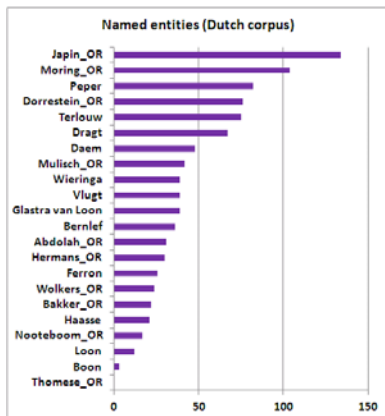


Fig. 5: all plot internal named entities in the Dutch corpus

[file: Personalnames_Karina_fig_5.png (colour)]

[file: Personalnames_Karina_fig_5_BLWH.png (black-and-white)]

The four books for children and young adults take ranks 5, 6, 10, and 21. The Dorrestein novel is, at rank 4, close to two of the children’s books, but this cannot be seen as significant until a lot more novels have been analyzed in the same way.

7 Conclusions

Several things have come to light in this contribution. Based on the measurements on a relatively small corpus of 22 Dutch and 22 English novels, we found that in this corpus the normal range of name tokens lies between 2 % and 3.5 %. Most of these name tokens are personal names; geographical names and other names play a relatively insignificant role, statistically speaking. Most of the personal names are plot internal names, so names referring to fictional characters. Most of the plot internal personal names are first names. And when we rank the 44 novels from the highest frequency to the lowest of name tokens, personal name tokens, plot internal personal name tokens, every time 6 of the 8 books for children and young adults end up at the top of the list. The ranking for plot internal first name tokens shows some relatively minor and easy to explain changes. Only when we look at the number of different plot internal *characters*, entities, that get a name,

the books for children and young adults do *not* cluster together at the top (they even do not cluster lower down in the list).

As to name functions, we saw how sparse name use does not mean that a novel is unreadable in any way. The lack of a name foregrounds certain themes and motifs in a novel, such as that names (and presumably identity or individuality) become useless or even dangerous in extremely horrible situations, as is even stated in quite explicit meta-onomastic phrases. We also saw how a higher frequency of personal names can function to highlight the complexity and oppressiveness of family and family life. These functional interpretations, however, still need to be more firmly proven, possibly with other visualization tools such as name density or nearness analysis.

We found that novels can do with only very little names or no names at all, but it still remains to be seen whether there is a maximum. Is it possible to write a novel that contains so many names that it actually becomes unreadable? We could think of the reaction of many people to certain Russian novels, stating that they contain too many names for the reader to nicely keep track of everybody. The question then is whether it is indeed the amount of names which may confuse readers, or perhaps the amount of named entities (characters), or even the number of different names being used for the same characters. This would be an extremely interesting topic for further research.

8 Coda

I will end this contribution going back to the Portuguese author José Saramago, who wrote the 'nameless' novel *Blindness* that was briefly described in the introduction. In all of his novels, Saramago shows special attention to names, writing about their importance in meta-onomastic statements in his novels – but hardly using them. One of his other novels, titled *All the Names*, ironically has only one name in it, namely that of the main character, Senhor José (also the first name of the author...). He is fifty years old and works as a low-level clerk in the Central Registry of Births, Marriages and Deaths in an again unnamed city. He lives in an old lean-to to the Registry building and in the evenings, using his key to the long unused inner door to the Registry, he starts to collect information about 100 famous people from the city where he lives. He collects all their individual names on special cards. When he starts working on the slightly less famous, he accidentally takes the record card of a 36-year old woman who is not famous at all. He realizes that she is much more important than all of his 100 famous people taken together. Obsessed, he goes in search of her but in such a helpless way that he is almost sure not to find her. When he then stumbles onto her death certificate, he goes to the Central Cemetery to find out more.

the Central Cemetery's unwritten motto is All the Names, although it should be said that, in fact, these three words fit the Central Registry like a glove, because it is there that all the names are to be found, both those of the dead and those of the living, whilst the cemetery, given its role as ultimate destination and ultimate depository, has to content itself only with the names of the dead. (Saramago 2000:188)

He finds the grave of the woman in the section for suicides.

Senhor José's boss has followed closely (and unknown by José) what he has been doing and he advises him what to do next: 'Make up a new card for this woman, the same as the old one, with all the correct information, but without a date for her death' (Saramago 2000:243). When Senhor José proceeds to do this, he is clearly executing what he already thought about when after finding the woman's grave he went to her apartment:

Here lived a woman who committed suicide for unknown reasons, who had been married and got divorced, who could have gone to live with her parents after the divorce, but had preferred to live alone, a woman who, like all women, was once a child and a girl, but who even then, in a certain indefinable way, was already the woman she was going to be, a mathematics teacher whose name while she was alive was in the Central Registry, along with the names of all the people alive in this city, a woman whose dead name returned to the living world because Senhor José went to rescue her from the dead world, just her name, not her, a clerk can only do so much. (Saramago 2000:236)

The importance of personal names, be it real or fictional, could not be expressed much better.

Bibliography

- Borin, Lars, & Kokkinakis, Dimitrios, 2010: Literary onomastics and language technology. In: Van Peer, Willie, Zyngier, Sonia, & Viana, Vander (Eds). *Literary Education and Digital Learning: Methods and Technologies for Humanities Studies*. IGI Global, 2010. Chapter 3. Pp. 53-78.
- Kohlheim, Rosa, & Kohlheim, Volker, 2011: Der literarische Name zur Jahrtausendwende: Andreas Maiers Roman 'Wäldchestag' als Beispiel. In: *Beiträge zur Namenforschung* N. F. 46. Pp. 269–285.
- Van Dalen-Oskam, Karina, 2005: Vergleichende literarische Onomastik. In: Brendler, A. und S. Brendler (Hrsg.). *Namenforschung morgen: Ideen, Perspektiven, Visionen*. Hamburg. Pp. 183–191. English translation (Comparative literary onomastics) available online at http://www.huygens.knaw.nl/wp-content/bestanden/pdf_vandalenoskam_2005_Comparative_Literary_Onomastics.pdf.
- Van Dalen-Oskam, Karina, 2006: Mapping the Onymic Landscape. In: Maria Giovanna Arcamone, Donatella Bremer, Davide de Camilli e Bruno Porcelli (eds.), *Il nome nel testo*. *Rivista internazionale di onomastica letteraria* VIII (2006). Atti del XXII Congresso Internazionale di Scienze Onomastiche, Pisa, 28 agosto – 4 settembre 2005 (Proceedings of the 22nd International Congress of Onomastic Sciences, Pisa 28. Aug. – 3. Sept. 2005), vol. III. Pp. 93–103.
- Van Dalen-Oskam, Karina (to be published): Immer nach Norden. Gebrauch und Funktion von Eigennamen im Roman *Oben ist es still* von Gerbrand Bakker. Ein

Pilotprojekt zur vergleichenden literarischen Onomastik. To be published in: Beiträge zur Namenforschung N. F.

Van Dalen-Oskam, Karina (under review): Names in novels: an experiment in computational stylistics.

Novels

Barrett, Andrea, 1993: The forms of water. London.

Dorrestein, Renate, 1996: Verborgen gebreken. Amsterdam/Antwerpen.

Dorrestein, Renate, 2003: A crying shame. Translated from the Dutch by Hester Velmans. London.

McCarthy, Cormac, 2006: The road. New York.

Saramago, José, 2000: All the names. Translated from the Portuguese by Margaret Jull Costa. London.

Saramago, José, 2005: Blindness. Translated from the Portuguese by Giovanni Pontiero. London.

¹ This contribution is a partly elaborated version of the paper ‘Personal names in literature: examples from recent novels written in Dutch and in English’ that was presented at the Ivar Modéer symposium in Uppsala in October 2011.

² The namelessness of *Blindness* is well-known to many readers, and for instance is also mentioned in the Wikipedia entry about the author: ‘In his novel *Blindness*, Saramago completely abandons the use of proper nouns, instead referring to characters simply by some unique characteristic, an example of his style reflecting the recurring themes of identity and meaning found throughout his work.’

http://en.wikipedia.org/wiki/Jos%C3%A9_Saramago, accessed 10 October 2011

³ This contribution builds on several earlier articles (Van Dalen-Oskam 2005, 2006) and two articles currently under review resp. in print: Van Dalen-Oskam under review and Van Dalen-Oskam to be published. References to more literary onomastic literature can be found there. The quantitative approach is quite new for literary onomastics; the only other scholars who, as far as I know, take this approach are Rosa and Volker Kohlheim, in Kohlheim & Kohlheim 2011.

⁴ The pipeline of named entity recognition tools for literary onomastics for Swedish literature as described by Borin & Kokkinakis (2010) has no parallel yet for Dutch literature. I am not aware of any such pipeline of tools developed for English literary studies either.