

## A literary rat race

Karina van Dalen-Oskam (karina.van.dalen@huygens.knaw.nl),

Huygens ING, The Netherlands / Universiteit van Amsterdam, The Netherlands

**Long paper presented at DH2016, 13 July 2016**

Abstract published online, [dh2016.org](http://dh2016.org)

### Introduction

'A hype already', was the Friday 28 January 2011 headline on the first page of the Dutch newspaper *NRCnext*. 'It's called the literary rat race K2, the simultaneous publication of Herman Koch's *Zomerhuis met zwembad* and Kluun's *Haantjes*'. The front-page story continues with literary critic Arjen Fortuin presenting an analysis of the two novels that were published the week before, both written by well-known Dutch authors and with the amazing first print runs of 80,000 (Kluun) and 100,000 (Koch) copies. These two novels are from totally different authors, who began their careers on opposite sides of the literary spectrum. However, Fortuin states, they seem to be converging. Koch started out as a 'literary' author not selling very well, but with his last book before *Zomerhuis met zwembad* (*Summerhouse with Swimming Pool*), *Het diner* (2009, translated into English as *The Dinner*, 2012) he turned to a wider audience, thus - according to literary critics - severely damaging his literary reputation. Kluun (the one-word alias of Raymond van de Klundert) started out as writing popular fiction with no literary pretensions at all with his much read but openly despised *Komt een vrouw bij de dokter* (2003, translated as *Love Life* (2007), under the name Ray Kluun). Quite unexpectedly, on 23 January 2011, Kluun's new novel *Haantjes* (which could be translated as 'Alpha-males') got a positive review from prominent literary critic Arjan Peters in *de Volkskrant*. Fortuin found this an additional reason to compare the two novels. His conclusion is: 'It's an uneven literary match – Kluun plays in a lower league – but the commercial battle of K2 could be a close tie – although here Koch also seems to have the best chances: the Alpha-males are in fact very light-weight'.

### The Riddle of Literary Quality

Both novels are on the list of 401 novels analyzed in the project *The Riddle of Literary Quality* (<http://literaryquality.huygens.knaw.nl/>). The aim of the project (running until 2017) is a stylistic analysis of novels in Dutch and to compare this analysis with readers' opinions. The corpus was based on a list of most sold and most lent titles in The Netherlands from 2010 to 2012, excluding titles first published in Dutch before 2007. It includes Dutch originals (such as the novels by Kluun and Koch) and translations. The list contains a lot of genre fiction such as thrillers, 'literary thrillers', and chick-lit, and

many titles the publishers categorized as 'literary novels', among which the K2 titles (Koch's *Zomerhuis met zwembad* and Kluun's *Haantjes*). In 2013, in an online survey titled *Het Nationale Lezersonderzoek* ('The National Reader Survey') we asked a wide audience of readers to indicate which of the 401 novels they had read, and for a smaller set of these novels how they rated them on two scales: one on the scale of literariness, from 1 (not literary at all) to 7 (highly literary) and one on the scale of general quality, from 1 (very bad) to 7 (very good). In total, 13,782 respondents completed the survey. By combining these data with an analysis of stylistic characteristics of the novels we expect we can discover which textual features may play a role in the current Dutch conventions of literariness. In this paper I will compare the K2 novels both in an exploratory analysis of the survey results on the scale of literariness and in stylometric analysis. Do the opinions of the K2 readers agree with the stylometric picture we get?

### **Zooming in on style**

'Style is a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively' (Herrmann et al. 2015). For the K2 case, I use the Stylo package in R (Eder et al. 2013). This R package is mainly used for authorship attribution. It compares texts based on the frequencies of a range of words or characters. Since words give more insight into the texts themselves than characters, Stylo can also be used for literary analysis beyond verifying authorship. In Fig. 1 I used Stylo to measure the distance between the ten 'literary' novels with the highest scores for literariness and the ten novels that got the lowest scores. The bootstrap consensus tree (a harmonization of cluster analyses) was based on the most frequent 100, 200, 300 etc. words (MFW) until 1000 MFW. There is a complete distinction between the "HIGH" and the "LOW" group.

**10 highest and 10 lowest scores for literariness in contrast  
Bootstrap Consensus Tree**

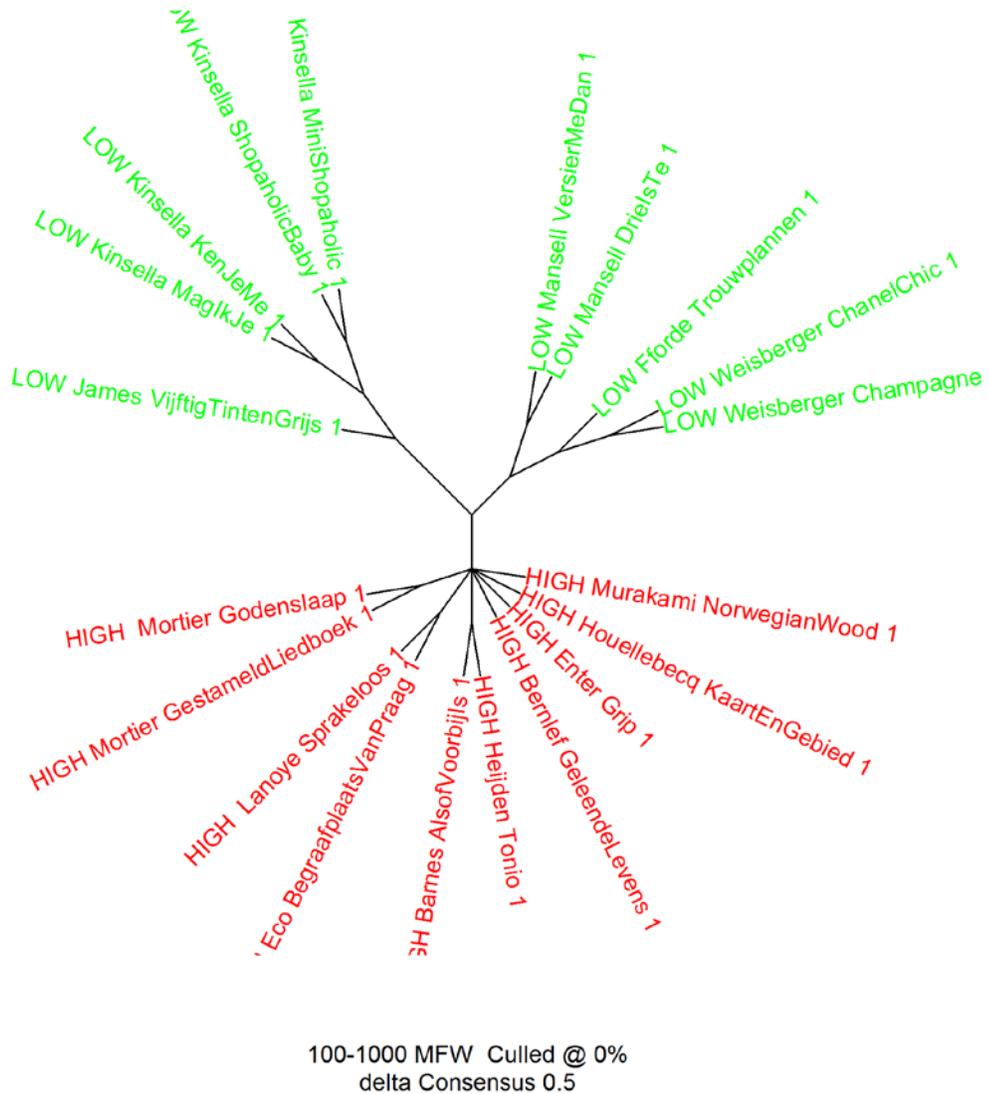


Fig. 1

However, the ten novels with the highest scores are all written by men and the ten with the lowest scores by women. The “HIGH” novels are labelled by the publishers as ‘literary novel’ and the “LOW” ones are mostly marketed as genre fiction such as chick-lit. The graph therefore probably does not distinguish literary quality but genre.

I now zoom in on the ‘literary novels’. The Riddle corpus contains 96 titles labeled as ‘literary novel’ written by a male author or by only male co-authors, and 66 by a female author or only female co-authors. For the whole set of 401 novels, 191 are written by male and 196 by female authors or co-authors, which shows female authors are not underrepresented in the corpus. In Fig. 2 all ‘literary novels’ are categorized according to the mean scores for level of literariness. The general trend seems to be that the respondents see the female authors ‘playing in a lower league’ than the male

authors. Riddle-PhD-students Corina Koolen and Kim Jautze will deal with gender issues in detail in their dissertations.

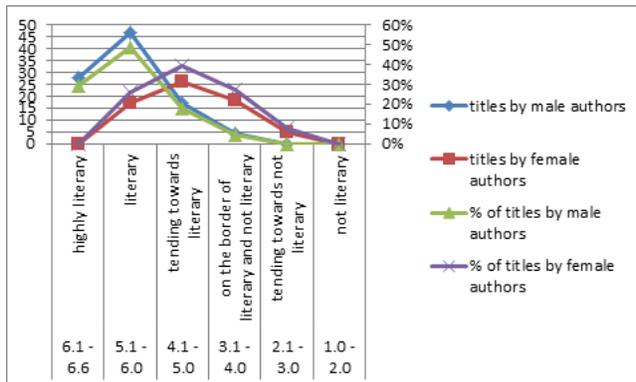


Fig. 2 Literary novels (originally Dutch and translated into Dutch)

I will leave the gender topic to my PhD-students to publish about, and I will for now limit my presentation to an analysis to novels written by male (co-)authors. The lowest score on the level of literariness was 3.1, for two titles written by the Dutch author (and sports reporter) Mart Smeets. Kluun's novel *Haantjes* is directly above these two novels, with a score of 3.5. *Zomerhuis met zwembad* clearly did better. It got a mean score of 5.1. The highest score is 6.6, for Julian Barnes' *Alsof het voorbij is* (*The sense of an ending*). Where Barnes' novel ranks first on the list of 96 novels, the new Koch ends up at rank 74 and the new Kluun at 94. The respondents of *The National Reader Survey* thus agree with Fortuin that Koch ranks higher on the scale of literariness. From the list of 96 novels I select thirty novels: ten with the highest (including Barnes), ten with the lowest (including Kluun), and ten with intermediate scores for literariness (including Koch). If we use the same settings in Stylo as above, the 30 novels end up as shown in Fig. 3.

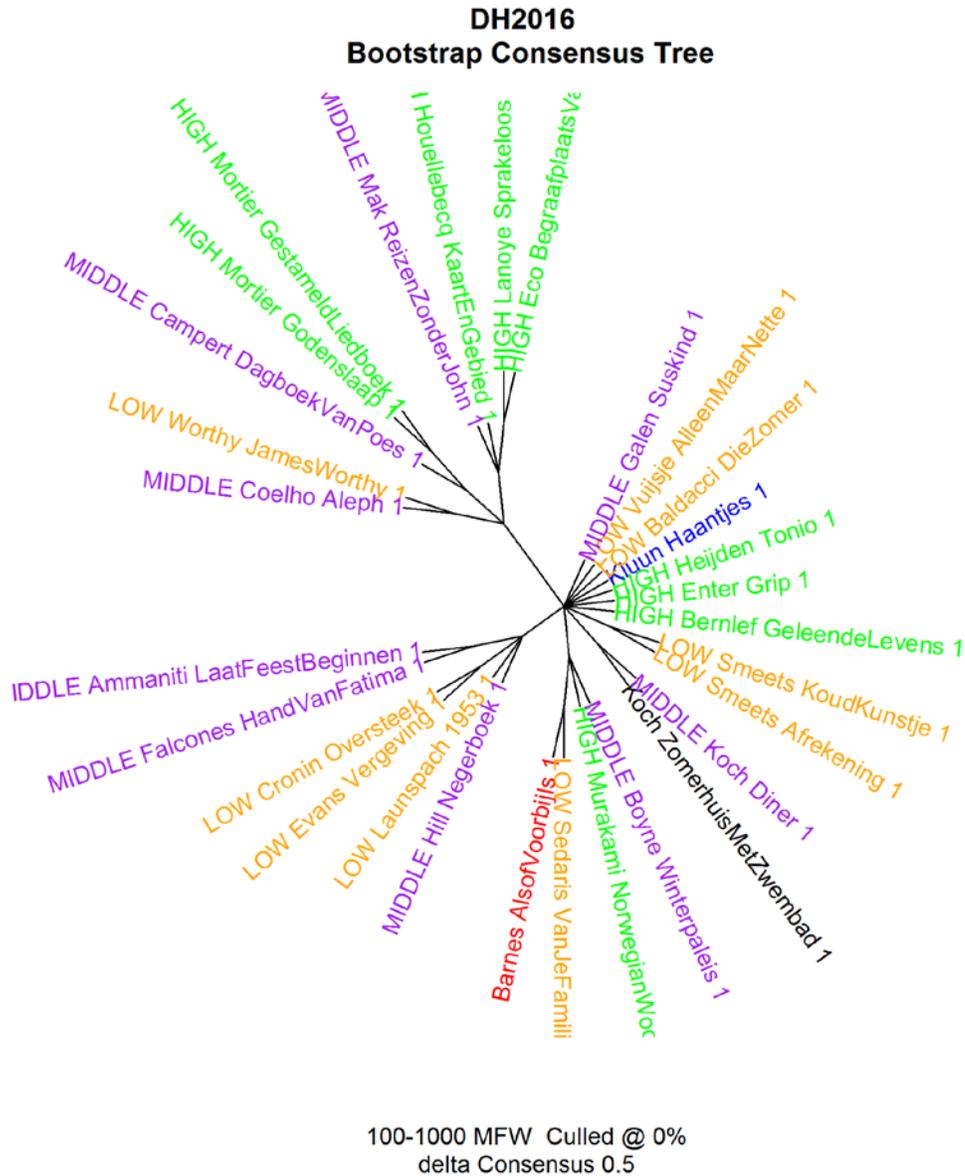


Fig. 3

The results are not very clear. The groups “HIGH”, “MIDDLE”, and “LOW” do not have their own specific clusters. For now, we can conclude that literary quality for this corpus does not reside in shared word frequency patterns. We are currently gathering as many measures of linguistic features as possible to apply these to the selected corpus. One of the assumptions to test is whether the scores for literariness correlate with features that relate to linguistic complexity. A suite of tools for Dutch is currently being developed. A simple test using HyperPo does show that some features normally related to the level of difficulty of a text need further inspection (Fig. 4, 5, 6) (<http://tapor1.mcmaster.ca/~sgs/HyperPo/>).

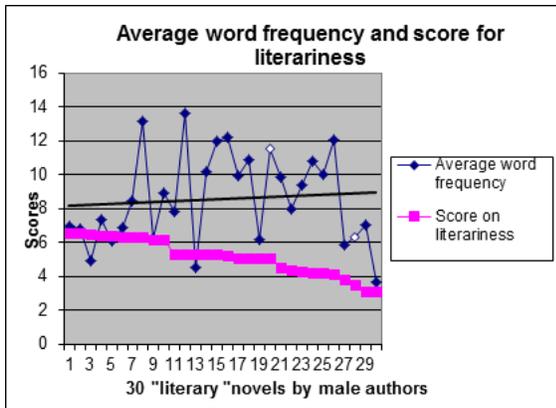


Fig. 4

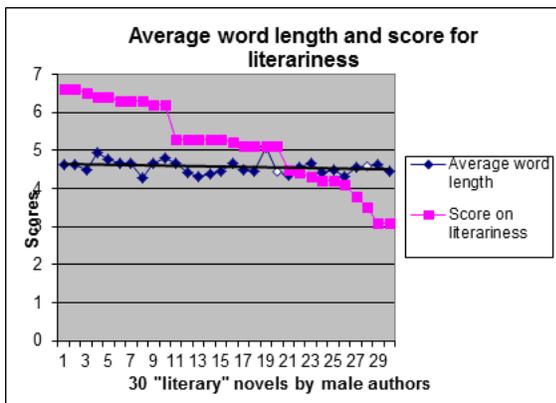


Fig. 5

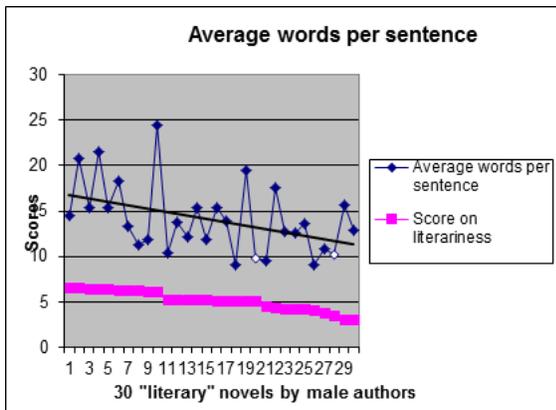


Fig. 6

In each of these figures, the 30 novels are represented on the x axis arranged from highest score for literariness to lowest. Koch is at data point 20 and Kluun at 28. Fig. 4 shows that the scores for literariness display a trend that is opposite to the average frequency of words – this suggests that for this corpus, lexical density scores perhaps play a role in what makes a novel literary or not. However, Kluun’s score below the

trend line is close to that of the top-5, while Koch is on the opposite side of the trend line. Fig. 5 highlights that mean word length does not resonate with scores for literary quality. Fig. 6, however, shows a clear (statistically significant) trend of literary score and average words per sentence. Here, Koch and Kluun do not differ very much. This sneak preview (which will be tested with tools finetuned for Dutch) shows a further analysis of complexity issues is promising. Riddle PhD student Andreas van Cranenburgh is working on syntactic markers of literariness, and all three PhD students have looked into topic variation in the Riddle corpus (see the abstract they have submitted for DH2016).

## Conclusion

So how do the novels by Kluun and Koch compare? The match is still undecided. We need an analysis of many more linguistic features before we can draw conclusions. This analysis did show in which directions to look. For now, stylistically they are playing in the same league.

## References

**Eder, M., Kestemont, M. and Rybicki, J.** (2013). Stylometry with R: a suite of tools. In: Digital Humanities 2013: Conference Abstracts. University of Nebraska--Lincoln, NE, pp. 487-89.

**Fortuin, A.** (2011) 'Mogen wij nog haantjes zijn? Tweestrijd van de moderne man is het gemeenschappelijke thema van Kluun en Koch'. [Can we still be alpha-males? Dilemma of the modern man is the shared theme of Kluun and Koch.] In: *ncrnext*, Friday 28 January 2011, p. 4-5. <http://www.nrc.nl/next/2011/01/28/mogen-wij-nog-haantjes-zijn-11991928>

**Herrmann et al.** (2015). J. Berenike Herrmann, Karina van Dalen-Oskam, Christof Schöch, Revisiting Style, a Key Concept in Literary Studies. *Journal of Literary Theory* 2015; 9(1): 25–52

**Kluun.** (2003) *Komt een vrouw bij de dokter*. Uitgeverij Podium, 2003

**Kluun, R.** (2007) *Love life*. Pan UK, 2007

**Kluun.** (2011) *Haantjes*. Amsterdam: Uitgeverij Podium, 2011

**Koch, H.** (2009) *Het diner*. Amsterdam: Anthos, 2009

**Koch, H.** (2011) *Zomerhuis met zwembad*. Amsterdam: Ambo|Anthos, 2011

**Koch, H.** (2012) *The dinner*. Faber & Faber, 2012

**Koch, H.** (2014) *Summerhouse with swimming pool*. Atlantic Books, 2014

**Peters, A.** (2011) '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 new Kluun is splendid. High quality entertainment in *Haantjes* by Kluun. Or: how Stijn van Diepen expected to make big money at the Gay Games in Amsterdam.] In: *de Volkskrant* 23 January 2011, <http://www.volkskrant.nl/boeken/de-nieuwe-kluun-mag-er-wezen~a1827236/>