

# The Literary Surface: A Plea for the Use of Measurement in Narratological Analysis

*Kim Jautze*

Huygens ING

kim.jautze@huygens.knaw.nl

*Roel Smeets*

Utrecht University

R.J.H.Smeets@uu.nl

Generally, analyses of literary representation are pursued on the basis of close readings of a confined amount of texts. In this mode of analysis, central narratological concepts as focalization are applied to literary cases in order to gain insight into (ideological) structures of representation (e.g. Song 2015; Minnaard 2010). Studies on, for instance, gender hierarchies in literature utilize Bal's (2009) distinction between a focalizing subject and a focalized object (e.g. Buikema 2009, Meijer 1996). However, the inevitable intuitive and subjective nature of close readings obstructs the generalizability of such interpretations.

In order to achieve a more intersubjective and generalizable interpretation, the computer can provide repeatable and verifiable methods to examine a large corpus of texts, thereby generating a more empirically solid basis. Digital tools, Moretti (2013) argues, may provide us with the opportunity to gain new insights into ever existing classifications in literary studies. Narratological concepts and hypotheses are being examined in the field of computational narratology. Whereas studies on superficial textual features (e.g. stylistic elements as the use of words or bigrams) do not necessarily require in-depth computational models, analysis of narrative features do require such models. In order to automatically detect character types without any prior knowledge requires building intelligent models (Bamman et al. 2014). For such studies it is necessary to have had a solid education in computer science, which most of the traditionally educated literary scholars do not have. It is therefore imaginable that they "fear" the rise of computational literary analysis. Underlying such resistance towards Digital Humanities could be 1) the difficulty of applying distant-reading approaches without being educated in programming and statistics, as well as 2) a general criticism on the reliability of computational measurements.<sup>1</sup>

Elaborating on that, we address two questions: 1) are there "low-level" textual features, i.e. features which lie on the texts' surface,<sup>2</sup> which can be examined with a basal knowledge of computers? And if so, 2) how reliable are the outcomes of such analyses? On the basis of the answers to these

---

<sup>1</sup> In this paper we do not address the arguments relating to the concern how literary scholars in Digital Humanities aim to come to objective measurements, which, in the opinion of most traditionalist, is in contradiction with the subjective nature of literary interpretations.

<sup>2</sup> Best and Marcus (2009) define the surface as "what is evident, perceptible, apprehensible in texts [...]. A surface is what insists on being looked at rather than what we must train ourselves to see through" (p.9).

questions, we explore to what extent a computational measurement of solely the texts' surfaces can contribute to, on the one hand, questions we ask our corpus with respect to narrative representations, and on the other hand, the already existing work on literary representation within the field of the "humanities narratology".

In our talk we will demonstrate how three relatively easy to use off-the-shelf tools do in fact complement a more traditional approach, and provide us - with high accuracy - the data we are in search of. In the first part, we apply a surface reading on a corpus of 45 contemporary Dutch novels;<sup>3</sup> that is, globally reading each chapter in order to identify characteristics of the narrators and main characters in the narratives. Secondly, we use AntConc and Stylo to analyse these particular findings in more detail. Because we are interested in the representation of the focalizers and narrators and their masculine or feminine 'voice' in the novels, we apply these tools in search of their share. More concretely, we count for an instance of the number of times the main characters are subject of an activity of perception. We make use of regular expressions with a predefined list of verba sentiendi adjacent to the subject of focalization, e.g. "John sees the treehouse". We do the same with a list of feminine and masculine pronouns. The main characteristic we are looking for is the gender of the narrators and main characters in order to measure the share between the male and female voices. Furthermore, the distribution of the surface characteristics can be compared among the 45 novels, for instance in search of the differences between male and female writers, between genres, or between more and less literary appraised novels. Thirdly, in order to demonstrate the merits and pitfalls of our method, we verify and test the computational measurements by manually tagging two cases from our corpus with CATMA. The results thereof make it possible to evaluate the accuracy of the tools applied. So far, our combination of a surface- and computational approach seems valid and accurate enough to provide a solid indication of the narrative surface features of our interest. For that reason, we argue, narratological research on literary representation could benefit from computational measurement of textual surface features.

## References

- Bal, Mieke. *Narratology: Introduction to the Theory of Narrative*. Toronto: 2009 (1985).
- Bamman, David, Underwood, Ted and Smith, Noah A. 'A Bayesian Mixed Effects Model of Literary Character. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. USA:2014, pp. 370-379.
- Best, Stephen & Marcus, Sharon. 'Surface Reading: An Introduction.' *Representations* (108):2009.
- Buikema, Rosemarie. *Doing gender in media, art and culture*. London:2009.
- Mani, Inderjeet. 'The Imagined Moment: Time, Narrative and Computation.' University of Nebraska:2010.
- Mani, Inderjeet. 'Computational Narratology'. *The living handbook of narratology*:2013 <http://www.lhn.uni-hamburg.de/article/computational-narratology> [last visited on 7th of January 2016].
- Meijer, Maaïke. *In tekst gevat: inleiding tot een kritiek van representatie*. Amsterdam:1996.
- Minnaard, Liesbeth. 'The Spectacle of an Intercultural Love Affair: Exoticism in Van Deyssel's *Blank en geel*'. In: *Journal of Dutch Literature* (1:1). 2010.
- Moretti, Franco. "'Operationalizing": or, the function of measurement in modern literary theory', *Pamphlets of the Stanford Literary Lab* (6):2013.
- Song, Angeline M.G. *A Postcolonial Woman's Encounter With Moses and Miriam*. New York: 2015.

---

<sup>3</sup> This paper is part of the project *The Riddle of Literary Quality*: <https://www.huygens.knaw.nl/projects/the-riddle-of-literary-quality/?lang=en>.