

The Riddle of Literary Quality Project

Kim Jautze*, Andreas van Cranenburgh*[†] and Corina Koolen[‡]

*Huygens Institute for Netherlands History, KNAW

[‡]Institute for Logic, Language and Computation, University of Amsterdam

[Kim.Jautze, Andreas.van.Cranenburgh]@huygens.knaw.nl; C.W.Koolen@uva.nl

Literary quality: a text-intrinsic property or determined by power? The formalists consider literariness to be a poetic function of language, caused by foregrounding elements that defamiliarize the reader, whereas literary sociologists argue that an author or his novel is accepted as literary depending on social and cultural factors. For the last decades this sociological approach of literary quality has been generally accepted in literary studies, resulting in a decline of research into intrinsic properties of texts such as their style (McDonald, 2007). However, textual and stylistic analysis are making a comeback. This phenomenon seems to coincide with the rise of the digital and computational methodologies within the humanities. Innovative tools and methods offer new opportunities of studying text-intrinsic properties of literature empirically. This is what we pursue in the Riddle project. We surmise that these intrinsic aspects may be of importance in the perception of aesthetic quality. At the same time we believe that sociological factors still influence reader's opinions about literariness.

In order to investigate the perception of quality we consider 400 modern Dutch novels and study both the readers' evaluations as well as formal characteristics of the novels. The goal is to find correlations of textual features with two sorts of appreciation, namely literariness and good versus bad. The aim is that our experiments and analyses can be verified and replicated by others. To this end we apply quantitative and computational methods on formal characteristics that can be measured objectively. Some of the features are directly observable in the text such as lexical richness or mean sentence length (we refer to these as low-level). Other characteristics, such as semantic, narrative or syntactic structure, require a deeper analysis of the text. Especially such high-level pattern recognition has hardly been applied to literary texts, neither for stylometry nor for studies on quality (Van Cranenburgh, 2012; Jautze et al., 2013). Therefore, most of the computational methods for these high-level patterns need to be developed.

In our talk we will first briefly summarize the set-up of the project. Then we will present some of the first results of the online survey *Het Nationale Lezersonderzoek* in which we have collected reader evaluations from 13.784 respondents. We will conclude with a couple of low- and high-level patterns we examined in the context of genre characterization.

References

Andreas van Cranenburgh (2012). 'Literary authorship attribution with phrase-structure fragments. In: *Proceedings of the Workshop on Computational Linguistics for Literature*, pp. 59–63.

Kim Jautze, Corina Koolen, Andreas van Cranenburgh & Hayco de Jong (2013), 'From high heels to weed attics: a syntactic investigation of chick lit and literature'. In: *Proceedings of the Workshop on Computational Linguistics for Literature*, pp. 72–81.

Rónán McDonald (2007), *Death of the Critic*. London / New York.

The Riddle of Literary Quality project: <http://literaryquality.huygens.knaw.nl/>