Suzanne Kleijn

Clozing in on readability

How linguistic features affect and predict text comprehension and on-line processing

The first readability formulae were developed almost 100 years ago. Despite a fair amount of critique, readability formulae have retained their overall popularity. The main reason for this is that the need for objective measures of readability has only increased. Fortunately, developments in computational linguistics have opened up new possibilities to improve the old readability formulae. In this dissertation current language technology is combined with insights from readability research and discourse processing in an attempt to build an empirically validated readability tool for Dutch secondary school readers.

We investigate the relationship between linguistic features and two aspects of readability: comprehension and processing ease. In addition, we use an integrated methodological design in which we combine experimental with correlational work to disentangle causal effects of linguistic features on readability from correlational relationships. That is, we study readability differences between texts and differences between stylistic variants of the same text. In three separate experiments we change only the lexical complexity, the syntactic complexity or the number of coherence markers within texts to see whether these factors really affect readability. This way we are able to provide a realistic (and sobering) view of the importance of these factors and their potential for reducing the difficulty level of a given text, without altering its content. Due to our design we are able to generalize our results across a large number of texts and across adolescent readers differing in reading proficiency. Hence, our findings are relevant both to the field of discourse processing and practitioners aiming for readability improvement.