Studies on automatic assessment of students’ reading ability

Katarina Heimann Mühlenbock², Erik Kanebrant¹, Sofie Johansson Kokkinakis²
Arne Jönsson¹, Caroline Liberg², Åsa af Geijerstam³, Johan Falkenjack¹
Jenny Wiksten Folkeryd³

Department of Computer and Information Science, Linköping University¹,
Department of Swedish, Göteborg University², Department of Education, Uppsala University³
katarina.heimann.muhlenbock@gu.se, arnjo@ida.liu.se

Abstract

We report results from ongoing research on developing sophisticated measures for assessing a student’s reading ability and a tool for the student and teacher to create a profile of this ability. In the project we will also investigate how these measures can be transformed to values on known criteria like vocabulary, grammatical fluency and so forth, and how these can be used to analyse texts. Such text criteria, sensitive to content, readability and genre in combination with the profile of a student’s reading ability will form the base to individually adapted texts. Techniques and tools will be developed for selecting suitable texts, automatic summarisation of texts and automatic transformation to easy-to-read Swedish.

1. Introduction

It is shown in different studies that even a not so strong reader is able to read in a more advanced way if the text is adapted with respect to aspects such as the topic of the text and different linguistic features, e.g. Liberg (2010), Reichenberg (2000). Our focus is to support reading for ten to fifteen year old students. The means for this is to find appropriate texts that are individually suitable and adapted to each student’s reading abilities.

2. Models of reading

Common to models of reading in an individual-psychological perspective is that reading consists of two components: comprehension and decoding, e.g. (Adams, 1990). Traditionally, the focus has been on decoding aspects, but in later years research with a focus on comprehension has increased rapidly.

The test of students’ reading ability in this study will include, in accordance with a broad view, different text types of different degrees of linguistic difficulty, where the students are tested for various reading practices within different topic areas. Items testing the following reading practices will be constructed for each of these texts, cf. Mullis et al. (2009, p. 23–29), OECD (2009, p. 34–44):

1. Retrieve explicitly stated information and make straightforward inferences (cf. text-meaning practices of Luke and Freebody (1999) and first envisionment of Langer (2011)),

2. interpret and integrate ideas and information (cf. Luke’s and Freebody’s text-meaning practices and Langer’s other envisionments), and

3. reflect on, examine and evaluate content, language, and textual elements (cf. Luke’s and Freebody’s pragmatic and critical practices).

Each of these practices also includes testing different aspects of vocabulary knowledge, c.f. Laufer and Nation (1995). These tests comprises everyday words originating from the same corpus as the readability texts.

3. Readability measures

We will consider global language measures built upon lexical, morpho-syntactic and syntactic features of a given text. The general readability of a text relates, however, not only to a combination of language properties making it easy or hard to grasp, but also to the specific reader (Mühlenbock and Johansson Kokkinakis, 2009). We will use the SVIT measures (Heimann Mühlenbock, 2013) which consider linguistic features appearing at the surface in terms of raw text, but also at deeper language levels. For the latter task we automatically process the text in four different steps: pre-processing, part-of-speech and lemma information annotation and finally parsing with dependency annotation.

The SVIT classification comprises four levels of linguistic complexity. The first level includes surface properties such as average word and sentence length and word token variation calculated with the OVIX formula. At the second level vocabulary properties are taken into account by measures of word lemma variation and the proportion of words belonging to a Swedish base vocabulary (Heimann Mühlenbock and Johansson Kokkinakis, 2012). The third, syntactic, level includes measuring the mean distance between items in the dependency trees, mean parse tree heights, the proportions of subordinate clauses, and pre- and post nominal modifiers. Finally, at the fourth level we considered the idea density present in the texts calculated in terms of average number of propositions, nominal ratio and noun/pronoun ratio.

A multivariate analysis revealed that for the task of discriminating between ordinary and easy-to-read children’s fiction, feature values at the vocabulary and idea density levels had the highest impact upon the positive results in automatic classification (Heimann Mühlenbock, 2013). For the present purpose, we therefore gave values indicating higher vocabulary diversity and difficulty metrics priority when the results did not unambiguously demonstrated any
A tool for teachers has been developed and distributed to all teachers with students that did the tests (Kanebrant, 2014). It allows teachers to get results on reading ability for each individual student. The tool is password protected to ensure that results only can be accessed by the teacher. The response texts intend to describe the readability competencies and vocabulary knowledge assessed. For the tests on reading ability we decided to group the categories 2 and 3 ending up with the two categories: 1) Retrieve explicitly stated information and make straightforward inferences and 2) Interpret and integrate ideas and information and reflect on, examine and evaluate content, language, and textual elements. We believe that it is easier for teachers to comprehend the results that way.

References