ABOUT KANSAS

Supports literacy teachers to find online reading material
• Flexible result re-ranking based on 85 linguistic constructions
• Visual enhancement of (de)prioritized constructions
• Filtering of results using our low literacy readability filter

Based on insights from SLA research to facilitate learning
• Supports literacy teachers to find online reading material

LOW LITERACY & FUNCTIONAL ILLITERACY IN GERMANY

Low literacy skills affect 13.3M people
• I.e. reading and writing below 9th grade level
• Causes: cognitive disabilities or disrupted education

Functional illiteracy affects 7.5M people
• I.e. cannot read or write short coherent texts

Degrees of (Il)literacy: Alpha Levels (Reikmann & Grotlüschen 2011)
• α1 reading/writing below character level
• α2 reading/writing at word level
• α3 reading/writing at sentence level
• α4 reading/writing of increasingly longer texts
• α56 reading/writing of short coherent texts

DE(PRIORITY & ENHANCEMENT OF CONSTRUCTIONS

Input salience & noticing are crucial for language acquisition
• Schmidt’s Notice Hypothesis (Schmidt 1990)
• Visual input enhancement (Shawcroft Smith 1993)

Facilitating noticing through (de)prioritizing and highlighting
• Identify 85 constructions on POS, phrase and sentence level
• Use Stanford CoreNLP API and Tregex for identification
• Re-rank using linguistic weights with BM25 IR algorithm

How well are we identifying constructions on web material?
• Pilot analysis of five constructions based on POS and Tregex

LOW LITERACY LEVEL CLASSIFICATION

Text Extractor

Web Search

Text Extraction

Linguistic Analysis

Alpha Readability Level Classification

Rule-Based Approach

FUNCTIONAL LITERACY

Search for challenging input within reach of learner (i+1)
• Krashen’s Input Hypothesis (Krashen 1977)
• Web queries mostly yield materials for advanced readers

Prioritizing linguistic constructions might decrease readability
• Easy materials may be unsuited for target audience

Ensuring readability through automatic filter
• Operationalized definitions from Alpha Levels to
• rule-based readability classifier
• annotation guidelines for humans
• Excluded α1 and α2 due to irrelevance for query results

Do our guidelines facilitate robust human judgments?
• Perform annotation experiment with two human raters
• 600+ texts from providers for low literacy reading materials

How well does our readability filter perform on (web) texts?
• Issue: no gold standard data set for evaluation available
• Create gold standard from annotation experiment

Pilot Evaluation
• Annotated 68 texts in pilot of annotation study with single rater
• Calculated IRR between algorithm and human rater

CONCLUSION & OUTLOOK

KANSAS assists teachers in finding suited materials
• Currently, teachers use search engines to find teaching material
• We promote the identification of texts that
• contain relevant linguistic constructions and
• have an appropriate reading level for low literacy readers
• Further uses include
• German as a Second Language learners
• web users with low literacy skills performing web queries
• The KANSAS search tool forms the basis of a collaboration with the DIE, Bonn and Mercator Institute, Cologne

Acknowledgements

KANSAS is a research and development project funded by the Federal Ministry of Education and Research (BMBF) as part of the AlphaDekade [grant number W143500].