# NLP System Demonstration

Xiaobin Chen

Tübingen University

May 17th, 2017
Flagstaff, AZ

# Common Text Analysis Platform

A Web-based tool supporting automatic analysis of text.

- Four components:
  - corpus manager
  - feature selector
  - analysis generator
  - result visualizer
- Features:
  - Consistent, easy-to-use, friendly user interface
  - Modularized, reusable, and collaborative development of analysis components
  - Flexible corpus and feature management
- Use cases:
  - Complexity analysis / readability assessment
  - Authorship attribution
  - Plagiarism detection
  - ...

# System Demo—CTAP

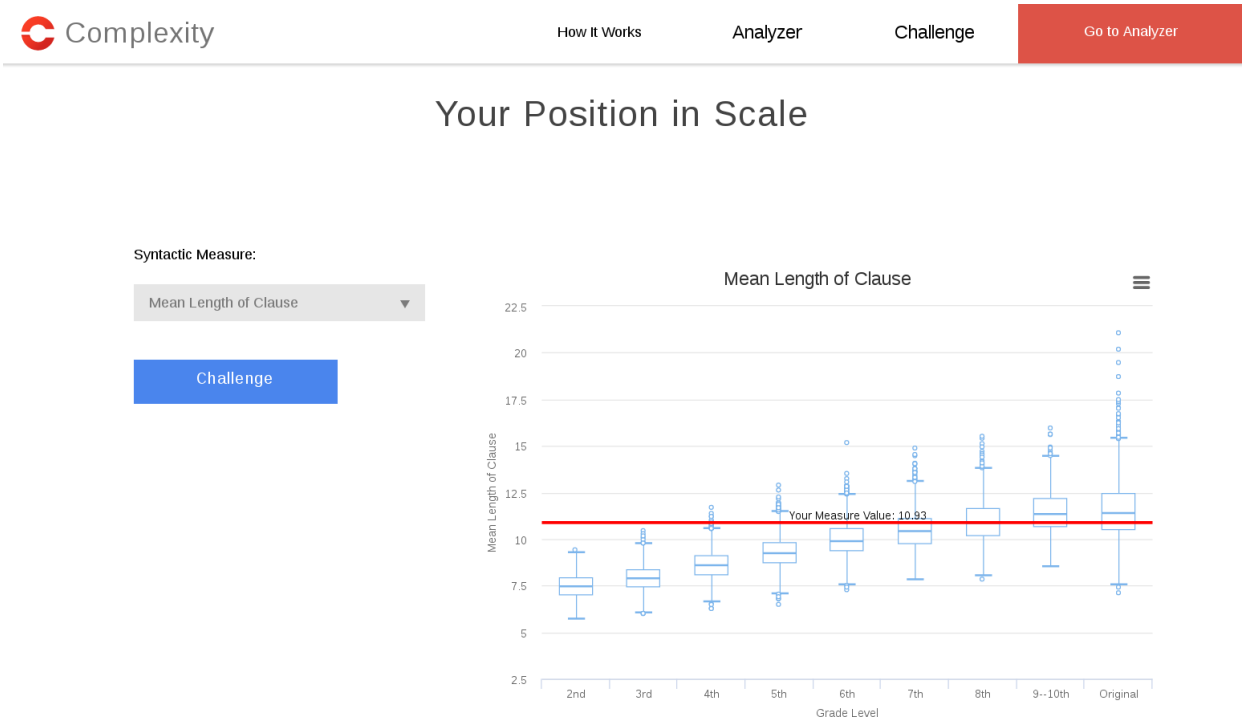

- `http://ctapweb.com/`
- Related publication:
  Chen, X.B., Meurers, D. (2016). CTAP: A Web-Based Tool Supporting Automatic Complexity Analysis. In *Proceedings of The Workshop on Computational Linguistics for Linguistic Complexity*. pages 113–119, Osaka, Japan, December 11–17 2016. The International Committee on Computational Linguistics.

# Syntactic Benchmarks

Challenges learners in their individual Zone of Proximal Development using pedagogic developmental benchmarks of syntactic complexity.

- Modeling syntactic complexity development with a target language or pedagogic corpus (Newsela, 14,581 news articles in 9 developmental levels)
- Placement of user proficiency on the developmental benchmark
- Provision of comprehensible reading input, which is configurable in terms of the degree of challenge and the target grade level
- Supporting 14 syntactic complexity measures

# System Demo—SyB



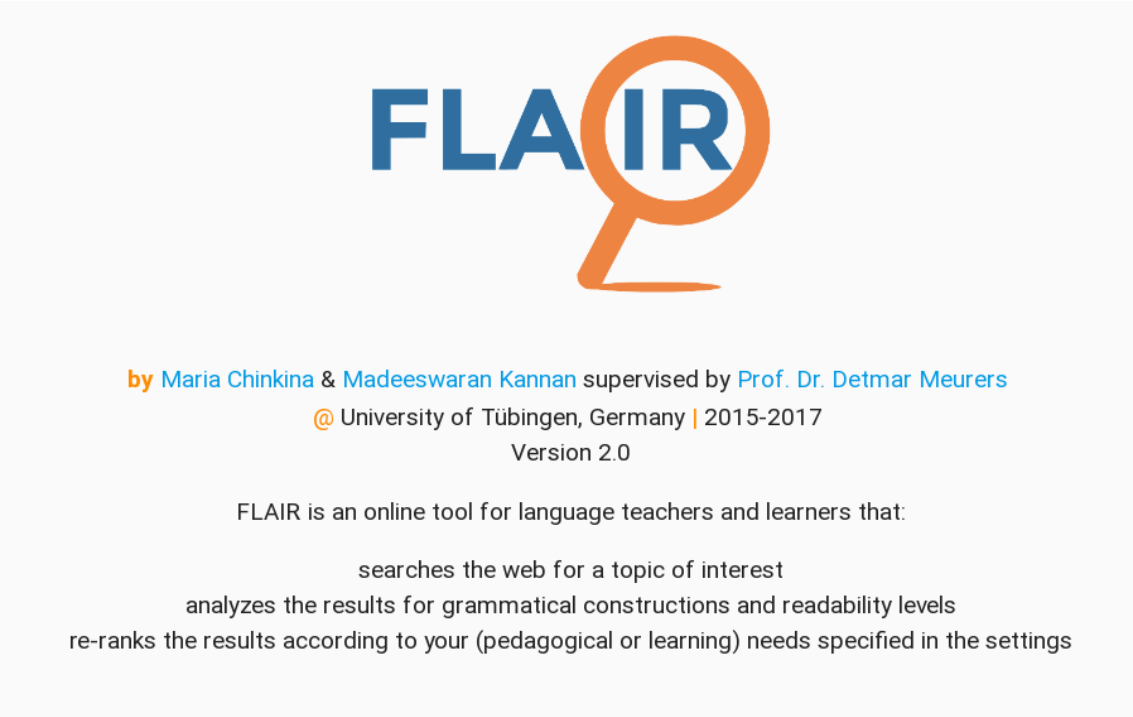

- Related Publication:
  Chen, X.B., Meurers, D. (2017). Challenging Learners in Their Individual Zone of Proximal Development Using Pedagogic Developmental Benchmarks of Syntactic Complexity. In Proceedings of the Natural Language Processing for Computer-Assisted Language Learning Workshop. Gothenburg, Sweden, 22 May.

# FLAIR

Form-Focused Linguistically Aware Information Retrieval

- Primary operations:
  - Web Search
  - Text Crawling
  - Parsing
  - Ranking
- Identifies the 87 grammatical constructions spelled out in the official English language curriculum of schools in Baden-Württemberg, Germany

# System Demo—FLAIR




`http://samos.sfs.uni-tuebingen.de:8080/flair-2.0/`

- Related publication:

  Chinkina, M., Kannan, Madeeswaran, and Meurers, D. (2016). Online Information Retrieval for Language Learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics—System Demonstrations*, pages 7–12, Berlin, Germany, August 7–12, 2016.

# NLP at Work?

- Which components/functions of these systems require NLP processing?
- What NLP processes are required to realize these functions?
- Think about your own teaching and research. How can the NLP techologies used in these systems be used to solve your problems?

# NLP at Work—An Example with the SyB

- Mean length of clause in tokens
- Formula: #tokens / #clauses
  - Sentence segmenter -> tokenizer -> count #tokens
  - Sentence segmenter -> tokenizer -> parser -> tree structure pattern matcher (Tregex) -> count # of matches of the clause pattern -> #clauses