NLP without Annotated Dataset

NLP and Language Learning

Sowmya Vajjala

Seminar für Sprachwissenschaft, University of Tübingen, Germany

29 January 2021

NLP and Language Learning

What are some applications?

Broadly, four categories:

- learning assessment
- learning support
- support tool for research (e.g., on language acquisition)
- learning analytics

Learning Assessment

Language Assessment

- Automatic essay scoring systems (such as e-rater by ETS)
- Automatic speech scoring systems (SpeechRater by ETS)
- Automatic creation of test items: multiple choice questions, fill in the blanks etc
- Active area of research for non-English languages too, in the past few years.

Content Assessment

- Evaluating short answers for correctness in relation to the question asked (i.e., going beyond looking for fluency, grammar/spelling correctness etc)
- To an extent, tougher than assessing language form.
- It was an active area of research a few years ago, and SfS had a strong group.

@SfS

In the SFB 833-A4 project, we are developing automatic meaning assessment methods for short-answer reading comprehension. To collect a rich task-based corpus in a real-life teaching context, we created the WELCOME app (Ott et al., 2012) and obtained the CREG corpus (36k answers to 1.5k questions). Our research showcases the importance of interpreting data in context (Ziai & Meurers, 2014; De Kuthy et al., 2015, 2016a, b; Ziai et al., 2016).

source: http://www.sfs.uni-tuebingen.de/icall/

NLP and Language Learning

What are some applications?

Broadly, four categories:

- language and content assessment
- **learning support**
- support tool for research on language acquisition
- learning analytics

Learning support

- Reading support: TextEvaluator (ETS) like tools to choose texts appropriate for a reading level
- Writing support:Spelling/grammar check tools (e.g., grammarly); specialized writing support (e.g., WritingMentor from ETS for academic writing support)
- Language learning apps (e.g., duolingo)

@SfS

- "Portuguese Intelligent Tutoring System (ITS) TAGARELA (Amaral Meurers 2011) designed to complement university instruction"
- "in collaboration with a German school book publisher we created the FeedBook, an interactive workbook for English 7th grade in a DFG-funded transfer project."
- "In the new BMBF project AISLA, we develop an intelligent dialog system supporting the acquisition of English in authentic, spoken language contexts. "
- "developing Prosodiya, a mobile serious game for German dyslexic primary-school children currently"

source: http://www.sfs.uni-tuebingen.de/icall/

@SfS-2

We are developing linguistic complexity analyzers integrating a wide range of linguistic, psycholinguistic, and SLA complexity features for English (Vajjala & Meurers 12, 13, 14a, b, c, Chen & Meurers 2016a, b) and German (Hancke, Vajjala, Meurers 12; Hancke & Meurers 2013) — and tools such as CTAP making it easy to use these measures.

Applying these methods to education, we investigate the (in)appropriateness of textbooks for students of different grades and school types (Bryant et al. 2017, Berendes et al., in press).

To support teachers and learners in identifying texts that are both interesting and richly represent the language constructs to be acquired, we created the linguistically-aware search engine FLAIR (Chinkina & Meurers 16). On this basis, we collaborate in the BMBF-funded KANSAS project with the German Institute for Adult Education (DIE) and the Mercator Institute for Literacy and Language Education to build a tool supporting teachers of functional literacy courses.

Connecting foundational and applied issues, we are spelling out Krashen's i+1 input fostering learning in terms of linguistic complexity using SyB (Chen & Meurers 17), a syntactic benchmarking tool, and we investigate the impact of challenging learners with such input.

CTAP | SyB | FLAIR | KANSAS

source: http://www.sfs.uni-tuebingen.de/icall/

NLP and Language Learning

What are some applications?

Broadly, four categories:

- learning assessment
- learning support
- **support tool for research**
- learning analytics

Language Acquisition Research

- Using NLP tools to study specific linguistic phenomenon in large learner corpora, to understand language acquisition
- Dependency parsing of learner language
- A recent paper: "Subcategorization frame identification for learner English"

etc.

@SfS

Studying learner language:

- "With Katrin Wisniewski we explored linguistic correlates of the CEFR as part of the MERLIN project."
- "We characterize language development both for specific constructions, e.g., relative clauses (Alexopoulou, Geertzen, Korhonen & Meurers, 2015) and in terms of linguistic complexity, emphasizing the need to account for task effects (Alexopoulou, Michel, Murakami & Meurers, 2017)."
- "We also analyze L1 transfer effects using machine learning for Native Language Identification"

source: http://www.sfs.uni-tuebingen.de/icall/

NLP and Language Learning

What are some applications?

Broadly, four categories:

- language and content assessment
- learning support for reading, writing, speaking, and listening
- support tool for research on language acquisition, learner corpora etc.
- **student data analytics**

Student data analytics

- modeling student engagement through their activity (incl. postings etc)
- summarizing course feedback given by students
- visualization of learning program etc.
- recent work from SfS: "Enhancing a Web-based Language Tutoring System with Learning Analytics"

Summary

- NLP is used in a wide range of topics related to human language learning.
- From research to industry, there are many interesting problems to study and solve.
- At SfS, there is a strong group focusing on this kind of research - so talk to them!

Where to look to know more

- BEA workshop series
- NLP4CALL workshop series
- Talk to Prof. Meurers and team.
- Two summary articles (by prof and ex-advisee):
  - Detmar Meurers (2013, 2020). Natural Language Processing and Language Learning. The Encyclopedia of Applied Linguistics, edited by Carol A. Chapelle. Wiley.
  - Sowmya Vajjala (2018). Machine Learning in Applied Linguistics. The Encyclopedia of Applied Linguistics (ed: Carol Chapelle). Wiley. (Okay, I am not sharing out of vanity. It really gives an overview)