LING 520: Computational Analysis of English

Semester: FALL '16

Instructor: Sowmya Vajjala

Iowa State University, USA

17 November 2016

Class Outline

- Assignment 5 Discussion
- Announcements
- NLP for CALL: overview of topics and research
- Update on final project status

Assignment 5 Discussion

Few announcements

- Grade improvement option: Do any of the following problems in Problem sets: 2.1-2.10, 3.1-3.6, 5.3, 7.5-7.10. Each problem carries 1 mark, and you can get a maximum of 7 points improvement. Send your code as a zip file, in an email, along with a pdf file listing what problems you solved from what problem set. Deadline: 8th December.

Few announcements

- Grade improvement option: Do any of the following problems in Problem sets: 2.1-2.10, 3.1-3.6, 5.3, 7.5-7.10. Each problem carries 1 mark, and you can get a maximum of 7 points improvement. Send your code as a zip file, in an email, along with a pdf file listing what problems you solved from what problem set. Deadline: 8th December.
- Class on 29th: Evgeny will give a talk about CyWrite. Read the following paper before attending the class:
  Feng, H. H., Saricaoglu, A., & Chukharev-Hudilainen, E. (2016). Automated Error Detection for Developing Grammar Proficiency of ESL Learners. CALICO Journal, 33, 49-70.
  (paper is uploaded on Blackboard)

Few announcements

- Grade improvement option: Do any of the following problems in Problem sets: 2.1-2.10, 3.1-3.6, 5.3, 7.5-7.10. Each problem carries 1 mark, and you can get a maximum of 7 points improvement. Send your code as a zip file, in an email, along with a pdf file listing what problems you solved from what problem set. Deadline: 8th December.
- Class on 29th: Evgeny will give a talk about CyWrite. Read the following paper before attending the class:
  Feng, H. H., Saricaoglu, A., & Chukharev-Hudilainen, E. (2016). Automated Error Detection for Developing Grammar Proficiency of ESL Learners. CALICO Journal, 33, 49-70.
  (paper is uploaded on Blackboard)

NLP for CALL - summaries of readings

- Idée on Meurers (2012)
- Liberato and Evan on Burstein (2009)
- Lena on Chapelle and Chung (2010)

Where is NLP useful for CALL?

There are two broad uses of NLP in the context of CALL.

- Analysing the language of learners (feedback, assessment, annotations etc)
- Analysing the non-learner language, for learners (suggesting materials, creating tests etc)

NLP for analyzing learner language

NLP for CALL: analyzing learner language

- Automated scoring of learner essays (e.g., in GRE exam)
- in language tutoring systems (e.g., TAGARELA, described in language and computers textbook) to create activities, give feedback, model a learner's error patterns etc.
- in offering on-the-fly feedback on writing (e.g., CyWrite, Grammarly, etc)
- in analysis and annotation of learner corpora (e.g., MERLIN http://merlin-platform.eu/)
- (often handled seperately): Analysing and assessing spoken responses of language learners

NLP beyond "language form" but within "learner language"

- Intelligent tutoring systems to teach science, psychology etc (holding dialogues with users, analysing their responses)
- systems like RWT which provide writing feedback, but more about content, and not about grammatical form
- scoring short responses of learners to questions about something they read

On using NLP for analyzing learner language

From Meurers (2012), page 2:

"However, there is an important difference in the goal of the NLP use in an ILTS compared to that in other NLP domains. NLP is made robust to gloss over errors and unexpected aspects of the system input with the goal of producing some result, such as a syntactic analysis returned by a parser, or a translation provided by a machine translation system. The traditional goal of the NLP in an ILTS, on the other hand, is to identify the characteristics of learner language and in which way the learner responses diverge from the expected targets in order to provide feedback to the learner. So errors here are the goal of the abstraction performed by the NLP, not something to be glossed over by robustness of processing."

On writers aids and learner language

From Meurers (2012), page 2

"Writer's aids such as the standard spell and grammar checkers (Dickinson, 2006) share the ILTS focus on identifying errors, but they rely on assumptions about typical errors made by native speakers which do not carry over to language learners. For example, Rimrott & Heift (2008) observe that "in contrast to most misspellings by native writers, many L2 misspellings are multiple-edit errors and are thus not corrected by a spell checker designed for native writers." Tschichold (1999) also points out that traditional writer's aids are not necessarily helpful for language learners since learners need more scaffolding than a list of alternatives from which to chose. Writer's aids tools targeting language learners, such as the ESL Assistant (Gamon et al., 2009), therefore provide more feedback and, e.g., concordance views of alternatives to support the language learner in understanding the alternatives and choosing the right. The goal of writer's aids is to support the second language user in writing a functional, well-formed text, not to support them in acquiring the language as is the goal of an ILTS."

NLP for analyzing language FOR learners

NLP for CALL: analyzing language FOR learners

- Selecting appropriate texts for learners (e.g., SourceRater/TextEvaluator by ETS: https://texteval-pilot.ets.org/TextEvaluator/)
- Modifying text (e.g., simplifying) for learners (e.g., Text Adoptor research prototype by ETS described in Burstein, 2009)
- Language Muse system -linguistically focused instructional authoring (ETS)
- Input enhancement - highlighting certain grammatical parts of a webpage to make learners notice (e.g., WERTI and FLAIRS, from University of Tuebingen)
- Creating multiple choice questions (or other forms of questions)

Useful resources to find readings

- CALL, ReCALL, CALICO etc journals in ALT community
- Educational Data Mining, AI in Education conferences and journals
- Building Educational Applications using NLP series of events
- International Speech Communication Association (ISCA) Special Interest Group: Speech and Language Technologies in Education (SLaTE)
- International Journal of Learner Corpus Research (not always computational)
- NLP4CALL workshop series in Sweden
- Natural Language Processing Techniques for Educational Application (NLP-TEA) organized by Chinese NLP community
- Shared tasks on grammatical error correction in ACL conferences

... and so on

Next Week

- NLP for CALL continued
- Evgeny's talk on tuesday. Read the article mentioned earlier in the class.
- We will have discussions and group presentations on all those those report writing questions in your assignments too.

Updates on final projects

- who is doing what?
- If you end up finishing your final project during the break, please volunteer to give a short presentation in the class in the last week. Do not consider this as extra work. Be proud of your project and be willing to show it off and get feedback.
- reminder about grading for the final project: 2.5 marks for initial report, 2.5 marks for final report. 10 marks for implementation. 5 marks for oral exam.

Happy Thanksgiving! Have a good break!