

LING 120, Fall 2017: Language and Computers

Instructor: Sowmya Vajjala

Iowa State University, USA

13 Sep 2017

Class outline

1. Recap of last class
2. Issues involved in generating practice exercises automatically
3. Linguistic analysis necessary to create such exercises
4. An example learning support system, and an exercise in understanding its working

Quick recap of last class

Computer Assisted Language Learning

Advantages:

- ▶ Individualized, personal feedback
- ▶ More and more practice with specific patterns of language, idioms of the language etc.
- ▶ No worries about limited classroom interaction with the instructor
- ▶ Unlike humans, a computer is objective with all students

Computer Assisted Language Learning

Advantages:

- ▶ Individualized, personal feedback
- ▶ More and more practice with specific patterns of language, idioms of the language etc.
- ▶ No worries about limited classroom interaction with the instructor
- ▶ Unlike humans, a computer is objective with all students

Questionable stuff:

- ▶ Can we trust a computer?
- ▶ Is it really possible to build such systems?
- ▶ Is it really free of bias?

Some tasks an intelligent CALL system should do

- ▶ Generate different kinds of questions
- ▶ Automatically evaluate answers
- ▶ Show different kinds of reading materials (from the web too, if needed - not limited to selected frames)
- ▶ Give feedback to learners if they make mistakes.

...

Yesterday's exercise

- ▶ Let us take this passage *The company was founded in 2009 by Alex Shevchenko and Max Lytvyn. Brad Hoover, the company's chief executive officer, is an investor with a background in engineering who learned about Grammarly while searching for an automated proofreading tool for his own writing. Grammarly, Inc, has headquarters in San Francisco. An additional office is in Kiev.*
- ▶ If I asked you to create multiple choice questions from this, what questions will you create? and why?
- ▶ Write down your questions on a paper and the choices, and the rationale behind choosing them.
- ▶ Work in groups of 2–3 people. Write your names on the paper and return it to me. This counts as your attendance for today.

Comments on Your answers

- ▶ There are several person names, so there can be a question related to names and writing combinations of names can create confusion while answering.
- ▶ There are several place names, so question can be asked about that.
- ▶ Question about year.
- ▶ Questions testing grammar (capitalization, missing apostrophe etc)

How can a computer create tutor and test automatically?

What kind of things should a computer know?

Reading and Writing

- ▶ break up sentences into individual words (called tokenization)
- ▶ Identify different parts-of-speech (to generate say questions that test prepositions), different forms a word etc
- ▶ Identify names of people, organizations, places, years etc (to generate factual, one word answer questions)
- ▶ Identify synonyms of a word, other related words such as hypernyms (Color is the hypernym of Red. Red is the hyponym of Color) etc. (Why?)

What kind of things should a computer know?

Reading and Writing

- ▶ break up sentences into individual words (called tokenization)
- ▶ Identify different parts-of-speech (to generate say questions that test prepositions), different forms a word etc
- ▶ Identify names of people, organizations, places, years etc (to generate factual, one word answer questions)
- ▶ Identify synonyms of a word, other related words such as hypernyms (Color is the hypernym of Red. Red is the hyponym of Color) etc. (Why?)

.. okay, may be all the others are difficult tasks. But, breaking up into words??

is tokenization even a problem?

- ▶ In the sentence: "Umm, I don't know it is perhaps a good idea to talk to a Dr. Vajjala instead of a laptop computer!", how many words are there?

is tokenization even a problem?

- ▶ In the sentence: "Umm, I don't know it is perhaps a good idea to talk to a Dr. Vajjala instead of a laptop computer!", how many words are there?
- ▶ Is Umm a word?
- ▶ Is "don't" one word or two?
- ▶ Does the dot after Dr mean a break into new sentence? or just a title?
- ▶ Should "laptop computer" be considered as one "compound" word? or two separate, unrelated words?
- ▶ Should we separate out those punctuations or keep them with words?

... a computer needs to concern itself about these issues.

What kind of things should a computer know?

Speaking and Listening

- ▶ break up speech into individual words
- ▶ Check for your pronunciation
- ▶ Other tasks can remain same here too.

How will it give back some feedback?

- ▶ It should understand what is missing in your answers.
- ▶ Let us say you used a wrong tense (i.e., it should know how to do grammar check)
- ▶ It should be able to generate a feedback sentence by itself and show/speak that to you.
("You missed a tense" can be a canned response. "In this sentence, you made a mistake with the tense of the verb pass" is a "generated" sentence)

Modeling the learner

- ▶ Keeping track of progress
- ▶ Understanding what a learner is doing well, what are her weak points
- ▶ over-use or under-use of specific language structures
- ▶ How to give personalized feedback (e.g., specific to the learner's native language background)
- ▶ sequencing of teaching materials
- ▶ prioritizing feedback (if there are several errors, should the system just keep putting red-marks all over, or prioritize one feedback over another?)
- ▶ plagiarism detection

.... and so on

An Example CALL system

- ▶ TAGARELA: Teaching Aid for Grammatical Awareness, Recognition and Enhancement of Linguistic Abilities
- ▶ a web-based workbook for teaching specific aspects of grammar.
- ▶ Website: <http://sifnos.sfs.uni-tuebingen.de/tagarela/index.py/main>
- ▶ Some aspects of TAGARELA (from a 2010 talk):
<http://www.sfs.uni-tuebingen.de/~dm/handouts/barcelona-10-10-26.pdf>

Other examples of tutoring/learning support systems

Next class!

Attendance exercise

post answers on Canvas forum for today

4. **ALL:** We mentioned multiple-choice exercises as a common CALL activity, for which all linguistic knowledge is directly encoded in the activity. Start a web browser and do a search for multiple choice language exercises. Explore the pages you find to see which language topics such multiple-choice exercises are used for. Why are those topics well suited to this exercise type?

Are there any differences between the exercises offered for learners of English and those for learners of Spanish (e.g., on the pages found by a search for Spanish multiple choice exercises)?

5. **ALL:** Many of the web pages identified by the search in the previous exercise also include exercise types other than multiple choice. Which kind of exercises do you find on those pages? Where do you think the knowledge to evaluate and provide feedback to those activities is encoded? Which kind of feedback do they provide? Are these canned responses?