LING 120:
Language and Computers
Semester: Fall '17

Instructor: Sowmya Vajjala

Iowa State University, USA

15 November 2017

Outline

- quick recap and Discussion about yesterday's exercise
- Group exercise + presentation on machine translation

Challenges for Machine Translation: Recap

- Each word can have multiple meanings, so mapping words one-one between languages is not possible.
- How words are formed between languages
- What is the order of words
- Words that don't seem to translate
- Words that seem to come up in translation that do not exist in original
- Issues such as: articles don't exist in Telugu, but they do in English. Kinship words are more fine grained in Telugu than English etc.
- Stylistic and cultural differences - which can make even manual translation difficult
- Metaphor is even difficult - same metaphors need not exist in both languages
- Poetry with rhyme - what if those words don't rhyme in Target language?

In this course so far, where is the idea of n-grams useful among the topics we discussed? Do you think they are useful in Machine Translation too?

Discussion about yesterday's exercise

- n-grams are useful for two aspects of machine translation
  - Map phrases in one language to another
  - Decide what is the most appropriate ordering of words in the translated output, what is grammatical, what is more natural for that language etc.

Rest of the class today

- A group exercise: Form into groups of 3 people (such that at least one person knows one non-English language)
- Two questions:
    - What other translation software exist apart from google translate?
    - Take a language pair (e.g., English-German, English-Chinese etc). Use google translate and try translating a few sentences. Do also the reverse translation. (Chinese-English, German-English etc.).
    - Try to understand how it works (e.g., where it works, where it fails, if I take an English sentence and translate to Chinese, and translate that Chinese one back to English, are they same? etc)
    - After 20 minutes, each team should summarize their observations - you can use 2–3 slides if you want (share the slides on Canvas forum for today and present to the class).
- This is your attendance.