C SC 620
Advanced Topics in Natural Language Processing

Lecture 19

4/6

Reading List

- *Readings in Machine Translation*, Eds. Nirenburg, S. *et al.* MIT Press 2003.

- Reading list:
  - 12. Correlational Analysis and Mechanical Translation. Ceccato, S.
  - 13. Automatic Translation: Some Theoretical Aspects and the Design of a Translation System. Kulagina, O. and I. Mel'cuk
  - 16. Automatic Translation and the Concept of Sublanguage. Lehrberger, J.
  - **17. The Proper Place of Men and Machines in Language Translation. Kay, M.**

Paper 17. The Proper Place of Men and Machine in Language Translation. M. Kay

- Introduction
  - Fully Automatic High Quality Translation (FAHQT) not realistic
  - Main point:
    - There is a great deal that computer scientists and linguists could contribute to the practical problem of producing translations, but, in their own interests as well as those of their customers, they should never be asked to provide an engineering solution to a problem that they only dimly understand
    - By doing only what can be done with absolute surety and reliability now and by going forward from there in short, carefully measured steps, very considerable gains can be virtually guaranteed to all concerned

Paper 17. The Proper Place of Men and Machine in Language Translation. M. Kay

- Machine Translation and Linguists
  - Translation for a pronoun
  - Target sentence pair:
    - Since the dictionary is constructed on the basis of the text that is being processed, **it** need refer to only a small amount of context to resolve ambiguities
    - Since the dictionary is constructed by a native speaker of the language, **he** need refer to only a small amount of context to resolve ambiguities
    - *Il* (French) -> {*it*, *he*}
  - Pronoun ambiguity
    - Ad hoc solution (case-by-case, statistical)
    - Not solved
  - Great number of ad hoc solutions built into existing MT systems and any enhancement in the future will require more and more of the same

Paper 17. The Proper Place of Men and Machine in Language Translation. M. Kay

- Machine Translation and Computer Science
  - Linguistic perspective: Unlikely to be adequate engineering where we know there is no adequate science
  - Programming style and techinque
    - Efficiency important - large quantities of text to be handled
    - Low level programming languages used (1980)
      - More likely to be ad hoc
      - Efficiency gains at most linear
    - To be continually improvable, program needs to be perspicuous (no doubt about the role that each of its parts play) and robust (can be changed in important ways without fear of damage)

Paper 17. The Proper Place of Men and Machine in Language Translation. M. Kay

- ## The Statistical Defense
  - Linguistics requires of its practitioners remarkable virtuosity in constructing examples of problems such as no existing or proposed computer system could possibly solve
  - The claim is that we do not have to solve them so long as they do not crop up very often
    - May not have an algorithm that will identify the antecedent of a pronoun whenever a human reader could, but if it can devise a method that will identify it most of the time, that will be good enough
  - Algorithm that works most of the time is of very little use unless there is some automatic way of deciding when it is and when it is not working
    - If algorithm could draw the proofreader's attention to all cases of pronominal reference that were in doubt, and if a high proportion of the cases were known to be correctly handled, then the utility of the technique is clear

Paper 17. The Proper Place of Men and Machine in Language Translation. M. Kay

- The Statistical Defense
  - If program works 99%, 90% , 80% or 50% of the time but we cannot tell which cases, the amount of work left to the repairman is essentially the same.
  - Real situation is worse: many decisions of the same difficulty must be made in the course of translating a single sentence. If there is reason to expect each of them to be correct 90% of the time, there need only be seven of them in a stretch of text to reduce the expectation to $.9^7=.48$
- The Sorcerer's-Apprentice Defense
  - Incomplete theories is a worse base to build systems than no theories
  - The man looked at the girl with the telescope
  - French admits parallel ambiguities in PP attachment
  - The man looked at the girl with penetrating eyes
  - *aux yeux* or *de ses yeux*, *avec yeux* not acceptable

Paper 17. The Proper Place of Men and Machine in Language Translation. M. Kay

- The Sorcerer's-Apprentice Defense
  - Ils signeront le document pourvu que leur gouvernement accepte
  - (I)
    - They {will, are going to} sign the document{ supplied, furnished, provided} that their government accepts
  - (II)
    - They {will, are going to} sign the document{provided that, on condition that, only if} that their government accepts
  - Intersect (I) and (II)
- The Translator's Amanuensis
  - transcriber
- Text Editing
- Translation Aids
  - Dictionary
  - Idioms (frequently occurring items)

Paper 17. The Proper Place of Men and Machine in Language Translation. M. Kay

- Machine Translation
  - Translator under the tight control of a human translator
  - Human/machine collaboration
  - Consult the user
  - It is far better that the labor and ingenuity spent on developing the machine's ability to make bad guesses should be employed more productively
  - Cascading errors common in machine-only systems
  - Keep track of words and phrases used in some special way

Readings

- ## Readings in Machine Translation
  - 19. Montague Grammar and Machine Translation. Landsbergen, J.
  - 20. Dialogue Translation vs. Text Translation – Interpretation Based Approach. Tsujii, J.-I. And M. Nagao
  - 21. Translation by Structural Correspondences. Kaplan, R. et al.
  - 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.
  - 31. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. Nagao, M.
  - 32. A Statistical Approach to Machine Translation. Brown, P. F. et al.

- ## Papers from recent MT Summit Conferences