NLP

Machine Translation

Advanced Methods

Other Techniques

- Tree-to-tree – Yamada and Knight
- Phrase-based – Och and Ney
- Syntax-based – Och et al.

Clause Restructuring (Collins et al)

- Ich werde Ihnen den Report aushaendigen … damit Sie den eventuell uebernehment koennen.
- I will pass_on to_you the report, so_that you can adopt that perhaps
- verb initial: that perhaps adopt can -> adopt that perhaps can
- verb second: so that you adopt…can -> so that you can adopt
- move subject: so that can you adopt -> so that you can adopt
- particles: we accept the presidency *Particle* -> we accept the presidency

(in German, split-prefix phrasal verbs are very common “anrufen” -> “rufen sie bitte noch einmal an” – call right back please)

Synchronous Grammars

- Generate parse trees in parallel in two languages using different rules
- E.g.,
  - NP -> ADJ N (in English)
  - NP -> N ADJ (in Spanish)

Machine Translation

Evaluation Methods

Evaluation

- Human judgments
  - adequacy
  - grammaticality
  - [expensive]
- Automatic methods
  - Edit cost (at the word, character, or minute level)
  - BLEU

BLEU (Papineni et al. 2002)

- Simple n-gram precision
- Multiple human references
- Brevity penalty
- Correlates with human assessments of automatic systems
- Doesn't correlate well when comparing human and automatic translations

Example from MTC

- http://www.eecs.umich.edu/~radev/nlp/mtc/
- Chinese:
  - Napster执行长希尔柏斯辞职
- English:
  - Napster CEO Hilbers Resigns
  - Napster CEO Hilbers resigned
  - Napster Chief Executive Hilbers Resigns
  - Napster CEO Konrad Hilbers resigns

Machine Translation

Decoding

Decoding

- Find a translation that maximizes P(F|E)P(E)
- NP-complete for IBM model 1
- Use a phrase translation table (e.g., Koehn's Pharaoh system, 2004)
- Use A* search to find the subset of phrase translations that covers the source sentence
- Combine with beam search

Tools for Machine Translation

- Language modeling toolkits
  - SRILM, CMULM
- Translation systems
  - Giza++, Moses
- Decoders
  - Pharaoh

NLP