

Assignment 6

Linguistics 445/515
Machine Translation & Dialogue Systems

Due Monday, November 24, 2008

1. Go to http://www.google.com/language_tools?hl=en. This online automatic translation system provides free translations between languages such as French, English, Spanish, Arabic, Chinese and Russian, among others. The language pairs marked BETA use statistical machine translation software developed by Google; the other language pairs use an older rule-based system.
 - (a) Try translating the following sentences from English into another language and then *back to English*. To do this, first translate from English to one of the other languages. Then, copy and paste the response and translate back to English. Try using different languages for each example. Write down the languages you translate to/from and what the backtranslations are.
 - i. It is a miracle that curiosity survives formal education.
 - ii. The spirit is willing, but the flesh is weak.
 - iii. These boots were made for walking, and that's just what they'll do.
 - (b) Rate the intelligibility of the translations: Is the word order correct? Does the system choose words that make sense? How close is the backtranslation to the original input sentence? Can you figure out the meaning or is it gibberish?
 - (c) Do some languages offer better backtranslations than others? For example, does an English-Spanish-English translation produce better results than English-Russian-English? Why or why not?
 - (d) Make up a sentence of your own and try it out. Write down both your sentence and its backtranslation. How well does Google do?
2. Take the following German and English sentence pairs and do the following (you might find a German-English dictionary helpful, although it is not strictly necessary):
 - (1) a. Das Kind ist mir ans Herz gewachsen.
I have grown fond of the child.
 - b. Sie ist bloß ein Kind.
She is but a child.
 - c. Sie nahm das Kind mit.
She took the baby with her.
 - (a) Describe precisely how the bag of words method derives a translation for *Kind*.

- (b) Describe how the expectation-maximization (EM) algorithm derives a translation for *but*. Address why we have to use several iterations when calculating alignments between words.
 - (c) How will phrase-based translation improve upon the translation models here?
3. Find a native speaker of a language other than yours (and other than English), and sit down with them with a short passage of text in their native language. Discuss what problems there are in translating from their language into English (or into your own native language). Which kinds of sentences/constructions are fairly easy to translate? Which ones border on impossible?
4. When translating from English into the Native American language Mam (in Guatemala), a translator reported the following terms used among siblings (in phonetic transcription here):
- [ntz?ica] = ‘older sibling’
 - [witzin] = ‘younger sibling’

Both words are used for males and females.

- (a) In terms of hyponymy/hypernymy, describe the relationship between the English word *sibling* and these words.
 - (b) Draw a Venn diagram showing how the English words *brother* and *sister* overlap with the Mam words *ntz?ica* and *witzin*.
 - (c) You come across the text: *Maxwell is the brother of Santiago*, but it gives no indication of who is older. If you had to translate this into Mam and were being forced to preserve this age ambiguity, how would you do it?
5. Read the following conversations and explain the Gricean maxim(s) that allow us to infer Janet’s meaning (note that more than one may apply for each situation):
- (a)
 - JACK: Do you think Gretchen is right for me?
 - JANET:¹ Well, she has a lovely smile, a great record collection, and I’ve never heard her sing off-key.
 - (b) While unloading groceries from the car, Janet accidentally drops a bag containing a dozen eggs and breaks them all. Jack hears the bag drop and starts the following exchange:
 - JACK: How are the eggs?
 - JANET: Well, the cookies seem to be okay.
6. I’m giving you the skeleton of a python program, available off the course webpage.
- (a) Write the function `match_or`
 - (b) Write the function `match_and`
 - (c) Let’s say the user is allowed to input queries of the form `X AND (Y OR Z)`, i.e., with a mixture of potentially embedded ANDs and ORs. Describe in prose (i.e., do not write python code) how you would break down these kinds of queries into a form that you could process and what necessary adjustments you would have to make to `match_and` and `match_or`

¹Janet does not think Gretchen is right for Jack.