

Linguistics 384

Homework 5

Machine Translation and Dialogue Systems

DUE: Wednesday, November 16, 2005

1. (20 points)

When translating from English into the Native American language Mam (spoken in Guatemala), a translator reported the following terms used among siblings:

- *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 (see slide 52) 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 were forced to preserve this age ambiguity in Mam, how would you do it?

2. (30 points) System evaluation exercise:

Go to: <http://www.tashian.com/multibabel/>

(This website is sometimes a little slow. Please be patient while waiting for the translation page to load.)

For the first three parts of this question, ignore the Chinese, Japanese, and Korean options.

- (a) Come up with an example sentence that you're going to translate and back-translate and write it down. Be funny, be creative, pick a song lyric or movie quote, whatever. Just make sure that the sentence is sufficiently interesting so that you are able to answer all of the following questions.
- (b) Enter your sentence, and examine all the (English) backtranslations. Write down all the backtranslations and for each backtranslation (there are 5 languages, so make sure you give me all 5 backtranslations), give me its score (1-4) on the intelligibility scale (given on slide 66).

- (c) In terms of quality, pick the best and worst backtranslations. Explain how you arrived at the best and worst – i.e. think about intelligibility, accuracy, error analysis. (For error analysis, think of criteria you can use for determining quality: meaning change, tense change [present, past, future], word choice, missing/added words, word order, “word salad,” etc.)
- (d) Now, turn on the Chinese, Japanese, and Korean option. Are these backtranslations generally better or worse than the others? WHY do you think that is?
3. (10 points) I’m trying to design an MT system so I can translate in and out of 4 different languages (Portuguese, Quechua, Romanian, and Slovenian), and I want to know how many *transfer* components I’m going to have to build if I: a) use a transformer system, and b) use an interlingua. Please show your work and explain your answer.
4. (20 points) <http://www.manifestation.com/neurotoys/eliza.php3> has a version of ELIZA you can play with online. (This is a slightly different version than we used in class.) Record an interaction you had with ELIZA (minimum of 5 interchanges – i.e. write down 5 sentences of yours with 5 ELIZA responses). Look at a snippet of the source code for this version of ELIZA by going to
- <http://www.ling.ohio-state.edu/~adriane/384/handouts/eliza.txt>
- (a) For each of ELIZA’s 5 responses, find the “response” line which generated it and give the response number (1-116).
- (b) In several of the responses, you can see “<*” at the end. What does this mean?
5. (20 points) Read the following conversations and explain the Gricean maxim(s) that Janet is using/flouting (note that more than one may apply for each situation):
- (a) Janet does not like Gretchen, but Jack may or may not know this:
- JACK: Do you think Gretchen is right for me?
 - JANET: 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.