

Homework 4

Linguistics 384 (Detmar Meurers)

Due at beginning of class on Tuesday, Nov 23, 2004

1. Go to <http://babelfish.altavista.com>, a site which allows you to type in text and translate it into another language. You can also *backtranslate*, i.e., translate back into the original language, by copying and pasting. To answer the following questions, you will have to come up with your own examples to defend your answers.
 - (a) Do you think this site uses a transformer system or a transfer system? Why?
 - (b) How deep do you think the grammatical analysis is, and why do you think so? I.e., Is the system using morphological analysis? Does it proceed word by word or is the whole sentence taken into account in a syntactic analysis? Is there evidence of a semantic or discourse analysis?

There is no specific right answer for this question, but I'm looking for evidence of thought in the argumentation. In particular, you should defend your answers with specific examples that illustrate your points.

2. (a) In the examples (1), (2), and (3) below, align the words in the English a. examples with the words in the Russian b. examples.¹
 - (1) a. This is a beautiful cat.
b. Eto krasivaya koshka.
This beautiful cat
 - (2) a. I have no money.
b. U menya net deneg.
By me no money
 - (3) a. I didn't know that I needed to go shopping for Eva.
b. Ya ne znal, chto mne nado bylo khodit' po magazinam Eve.
I not knew that to me need was to go around stores to Eva
- (b) How would you use the alignments in (a) to calculate probabilities of translations?
- (c) If you didn't have alignments, you could use a bag of words model. How would your probabilities for these examples be different from those you calculated under (b) ?
- (d) The bag of words model, of course, gets better over time. Describe how these extra sentences in (4) and (5) help you translate certain words. (i.e., Which words get easier to translate and why?)

¹Note that multiple English words may correspond to a single Russian word and vice versa, and some words might not align with anything. Note also that I have provided a word-by-word transliteration underneath the Russian—this is NOT what you are aligning with; this is only to let you know what each word in Russian roughly means. So, e.g., in (1), align a. “This is a beautiful cat.” with b. “Eto krasivaya koshka.”

- (4) a. I bought a cat.
 b. Ya kupil koshku.
I bought cat
- (5) a. Ivan thought that I knew.
 b. Ivan dumal, chto ya znal.
Ivan thought that I knew
- (e) Note how the Russian word for *cat* changes depending on how it's being used in Russian. What would you have to do to translate *cat* into Russian appropriately with the bag of words method?
3. Go to <http://www.tashian.com/multibabel/>
 For the first four parts of this question, ignore the Chinese, Japanese, and Korean options.
- (a) Come up with an example sentence (10 words or longer) that you're going to translate and backtranslate 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, give me its score (1-4) on the intelligibility scale (cf., our machine translation handout).
- (c) In terms of quality, pick the best and worst backtranslations. Explain how you arrived at the best and worst – i.e., provide an argument which takes into account 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, 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?
4. When translating from English into the Native American language Mam (in Guatemala), a translator reported the following:
- 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 the Mam words *ntz?ica* and *witzin*.
- (b) Draw a Venn diagram showing how the English words *brother* and *sister* overlap with the Mam words *ntz?ica* and *witzin*.
- (c) You want to translate the sentence *Maxwell is the brother of Santiago.* into Mam, but there is no indication of who is older. How would you translate the word *brother* in the sentence into Mam in a way that preserves this uncertainty about the ages?