

Assignment 6

Machine Translation Dialogue systems

DUE: Wednesday, February 25

1. (a) Align the words in the English (a) examples with the words in the Russian (b) examples. Note that multiple English words may correspond with the same Russian word and vice versa and that there might be null elements. 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 *This is a beautiful cat.* with *Eto krasivaya koshka.*
 - (1) a. This is a beautiful cat.
b. Eto krasivaya koshka.
This beautiful cat
 - (2) a. I have no money.
b. U menya net vremeni.
By me no money.
 - (3) a. I didn't know that I needed to go shopping for Natasha.
b. Ya ne znal, chto mne nado bylo khodit' po magazinam
I not knew that to me need was to go around stores
Natashe.
to Natasha
- (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 be different from part (b) for these examples?
- (d) The bag of words model, of course, gets better over time. Describe how these extra sentences help you translate certain words. (i.e. Which words get easier to translate and why?)
 - (4) Ya kupil koshku.
I bought cat
'I bought a cat.'
 - (5) Ivan dumal, chto ya znal.
Ivan thought that I knew
'Ivan thought that I knew.'

- (e) What would you have to do to translate *cat* into Russian appropriately with the bag of words method?
2. (a) Create a dictionary entry for an English to Italian MT dictionary for the words *knotty*. You may use the examples in class as a basis for this (p. 40-41)¹, but I suggest you add features where necessary (be thorough). Use www.wordreference.com. (You may use another source, but please indicate if you have.)
- (b) Take the Italian word(s) you used above and create a dictionary entry, or entries, for it/them for an Italian to English MT dictionary.
3. While roaming the Saharan desert, Professor Peradventure came across some ancient hieroglyphics which described a perfectly “logical” language, where everything you would ever want to express can be expressed in clear, unambiguous terms. For example, in (6), the action is first defined; then the object undergoing that action is given. And then – and this is the brilliant part – the object moves in a straight line, so to speak, from its origin (*me*) to its destination (*Shnupart*).

- (6) Mopar fingle lolo shnupart.
 Give ball-the from me to Shnupart
 ‘I give the ball to Shnupart.’

So, Professor Peradventure has this great idea (and he’s not the first) to use this language – named Shnupartese after its most famous speaker – as an *interlingua* in a machine translation system. So, he quickly puts together a Russian → Shnupartese → English system.²

- (a) Show and describe the steps that this MT system would go through to translate (7a) into (7b).
- (7) a. Ya podaryu Dzhony myach.
 I give to John the ball
 b. I am giving the ball to John.
- (b) I don’t want to break Professor Peradventure’s heart, but I’m not sure that Shnupartese is as logical as he thinks it is. Reorder the words in (6) and defend its logicalness.
- (c) Which of the 16 possible orderings of (6) would be easiest for a computer to understand?³
- (d) In English we say *I am giving the ball to John.*, but we also have another way of saying this. What is the other way? Does our interlingua have a way to distinguish these two sentences?
- (e) Give me a context where *I am giving the ball to John.* sounds better than your answer in part (d). (Try emphasizing different words and think of contexts that go with those emphases.)

¹Remember that *knowledge* should have been “CONCRETE: no” because knowledge is not something you can touch.

²Peradventure spent so much time in the Sahara that he is under the impression that the Cold War is still going on, so this is his obvious first choice.

³Is this a trick question, Mr. Dickinson?

(f) Give me a context where your answer from part (d) sounds better than *I am giving the ball to John*.

4. <http://www.stanford.edu/group/SHR/4-2/text/dialogues.html>

gives a conversation between ELIZA and PARRY. You have to go about halfway down the page to the section appropriately titled, “a conversation between eliza and parry”. (The rest of the webpage is fun to read, but this is the only part we will examine.) Point out 2 things that make this dialogue sound unnatural: give me the problematic parts of the dialogue and a brief (1-2 sentence) description of the unnaturalness.

5. <http://www.manifestation.com/neurotoys/eliza.php3> has a version of ELIZA you can play with. Record an interaction you had with ELIZA (minimum of 5 interchanges – i.e. 5 sentences of yours with 5 ELIZA responses).

Look at the Page Source code, beginning with the part which says, “// build our data base here” until “response[116]”

(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, we find “<*”. What does this mean?

6. Give me an outline for your own ELIZA-type system. The task for your ELIZA (or whatever you want to call him/her/it) is to be a hostage negotiator. I want you to sketch for me how you would begin to design such a system.

Issues to consider include:

- What kinds of input do you expect? (Are there certain templates to include?)
- What words/phrase (stock phrases) would you be sure to include?
- What words/phrases do you want to avoid?
- How would you handle demands? i.e. How would your ELIZA make decisions? How would your ELIZA know what to agree with?
- What do you see as the biggest problem for your system? (Lack of knowledge? Lack of memory? Lack of grammar? etc.)