COGS 200 Assignment 3 Brian Taylor

1a. Consider the following pairs of sentences:

(1) I wrote a letter to Bob. I wrote Bob a letter.

(2) Jane gave me some cake. Jane gave some cake to me.

(3) The teacher sent a package to her friend. The teacher sent her friend a package.

Describe one transformation that could relate the two sentences in the pairs. (2 points)

The following transformation works well for the three given sentences. The word 'to' is dropped and the positions of the direct object and indirect object are swapped. Sentences (1) and (3) start with the pattern on the left and get transformed to the pattern on the right, whereas sentence (2) goes in the reverse order.

<subj.> <verb> <direct obj.> to <indirect obj.>. → <subj.> <verb> <indirect obj.> <direct obj.>.

/2

1b. Now consider this pair of sentences:

(4a) He mailed a note to Germany.

(4b) ?He mailed Germany a note.

Does your transformation capture the intuition that (4b) is a little odd? If it does, how does it? If it doesn't, why not? (2 points)

The transformation I gave in part (a) has been applied to sentence (4a), but it leads to an awkward sentence that doesn't capture the meaning of the original sentence. The difference between sentences (4a/4b) and sentences (1) to (3) is that in the first three sentences the indirect object is a person (… to Bob, … to me, … to her friend), whereas the indirect object in sentence (4) is a place. We could view this as a form of ambiguity, since in English we use the preposition 'to' in both cases, but in other languages a different preposition, or formulation, is used in the two cases respectively.

A more specific transformation could be created, as follows:

If <indirect obj.> is a person:
<subj.> <verb> <direct obj.> to <indirect obj.>. → <subj.> <verb> <indirect obj.> <direct obj.>.

Else if <indirect obj.> is a place:
<subj.> <verb> <direct obj.> to <indirect obj.>. → <subj.> <verb> to <indirect obj.> <direct obj.>.

If the second transformation is applied to sentence (4a) this still leads to a slightly awkward sentence:

He mailed to Germany a note.

However, it's grammatical and does capture the original meaning of sentence (4a).

/2

1c. Finally, consider this pair of sentences:

(5a) Joe addressed the letter to Ed's home address.

(5b) *Joe addressed Ed's home address the letter.

Does your transformation capture the intuition that (5b) is "bad" or "ungrammatical"? If it does, how does it? If it doesn't, why not? (2 points)

The transformation I outlined in part (b) turns sentence (5a) into:

Joe addressed to Ed's home address the letter.

This is a slight improvement on sentence (5b), but really not very good. So, perhaps we could consider more rules for our transformation. The difference between sentences (4a/4b) and sentences (5a/5b) is that the word 'to' is once again being used in a slightly different way. The 'to' of sentence (5a) does not mean that the letter will be sent *to* Ed's home address. It merely means that the address will be placed *on* the letter. The modified transformation of part (b) will not do a good job of distinguishing this, however, as Ed's home address could be interpreted as a place. But if you think about it a bit more carefully you can see that while Ed's home is definitely a place, Ed's home address is merely a description of that place. So, perhaps it's not unreasonable after all to expect sentence (5a) to skip past both of our conditions, thus requiring a third condition for our transformation:

If <indirect obj.> is a person:
<subj.> <verb> <direct obj.> to <indirect obj.>. → <subj.> <verb> <indirect obj.> <direct obj.>.

Else if <indirect obj.> is a place:
<subj.> <verb> <direct obj.> to <indirect obj.>. → <subj.> <verb> to <indirect obj.> <direct obj.>.

Else if <indirect obj.> refers to a place:
<subj.> <verb> <direct obj.> to <indirect obj.>. → ???

The question becomes 'what is a suitable transformation?' It really doesn't work well to simply swap the two objects as in the first two transformations. We could re-write the sentence in other ways, passive voice for example:

Else if <indirect obj.> refers to a place:
<subj.> <verb> <direct obj.> to <indirect obj.>. → <direct obj.> was <verb> by <subj.> to <indirect obj.>.

Or even:

Else if <indirect obj.> refers to a place:
<subj.> <verb> <direct obj.> to <indirect obj.>. → <direct obj.> was <verb> to <indirect obj.> by <subj.>.

The first of these would result in (5a) becoming:

The letter was addressed by Joe to Ed's home address.

And the second of these would result in (5a) becoming:

The letter was addressed to Ed's home address by Joe.

Both versions would be grammatical and easily understood, but might displease certain grammar mavens.

1d. How would an exemplar approach capture (i) the relationship between the pairs of sentences in (1), (2), and (3)? (ii) the oddness of (4b)? and the "badness" of (5b)? (3 points)

In the exemplar approach words and structures are recorded, or learned. In other words, the conditions that I had to add to the first transformation would be learned as structural exemplars. The exemplar approach explains why native speakers of a language just 'know' that a sentence is wrong. It's wrong because it sounds wrong. I once lived with a Japanese housemate whose English was sufficiently expert that she was studying to become a teacher of English as a foreign language. She would often ask my opinions about various sentences or grammatical constructions. I wasn't much help to her beyond saying that a particular sentence sounded correct or definitely sounded incorrect. She wanted to know the rules so that she could explain the rules to her students. And, of course, the problem with developing rules, as this whole question has just demonstrated so well, is that as soon as we have a rule we have exceptions to that rule, or sub-rules or conditions for that rule to be applied. Developing all these rules may be valuable to linguists who want to compare languages, or study language acquisition, but they don't seem to fit the lived experience of speaking a native language. I am not sufficiently fluent in any second language to apply this idea to the learning of a language as an adult, although in my limited studies of French, German, Japanese, Cantonese and Italian I find that the instruction is usually a blend of the exemplar approach and rules. For example, in Italian it took me a while to figure out the conjugation of the verbs *piacere* and *dispiacere*. I finally realized that in Italian these verbs are always used in something like the passive voice. For example instead of a direct 'I like red wine.' the sentence in Italian would come out more like 'Red wine pleases me.' Once I learned that rule I was able to use *piacere* and *dispiacere* without difficulty. I doubt any native speaker of Italian ever really thinks much about it. They would simply have learned from an early age how to express their likes or dislikes and no rule would have been necessary. If pressed, they could probably articulate the rule I came up with, but I doubt that is present in their minds before being asked.

3

2a. Briefly describe a real world example of a COGS type project that depends upon the frequentist interpretation of probability. Include an appropriate reference. Make clear what about the approach makes the interpretation frequentist. (3 points)

Google Translate utilizes the frequentist interpretation of probability. In order to get a reasonably usable translation system for a new pair of languages using statistical machine learning Franz Och of Google suggests that a corpus of one billion words of bilingual text and one trillion words of monolingual text is needed.

Using this enormous database the translation program can examine the frequency distributions of all possible *n*-grams (groupings of *n* words) and then use that frequency distribution to assign a probability to the likelihood of a particular *n*-gram. A model can then be developed to combine the respective probabilities of *n*-grams in a set to the likelihood of various sentences using those *n*-grams.

(source: Franz Josef Och. Statistical Machine Translation: Foundations and Recent Advances. Tutorial at MT Summit 2005. Phuket, Thailand, accessed at: http://www.mt-archive.info/MTS-2005-Och.pdf)

3

2b. Briefly describe a real world example of a COGS type project that depends upon the Bayesian interpretation of probability. Include an appropriate reference. Make clear what about the approach makes the interpretation Bayesian. (3 points)

Many spam filters utilize a Bayesian interpretation of probability. Individual messages are assigned a probability of being spam by the spam filter. This initial probability could be rule based (e.g. the message contains the word 'Viagra') or it could be based on prior assessment by the spam filter. The receiver of the message has the option to identify the message as spam, or alternatively to look in his or her spam folder and identify messages as not spam. This 'evidence' can be used to create a new probability assessment based on Bayesian updating. This technique is known as Naïve Bayes Filtering.

(source: en.wikipedia.org/wiki/Naive_Bayes_spam_filtering)

/3

3. Suppose you heard someone say, "I saw the man with a telescope." How would an exemplar approach describe your ability to understand the ambiguity in that sentence? Could an exemplar approach resolve the ambiguity (that is, decide which meaning is intended by the speaker)? If so, how? If not, why not? (6 points)

The ambiguity of this sentence arises because the word telescope could be something the man has or it could be something we use to see the man. There would be no ambiguity if the object were a book instead of a telescope since we can't see with a book. I don't think the exemplar approach can resolve this ambiguity on its own. I know that when I write a sentence that could be interpreted in two ways I try to come up with a formulation that is unambiguous. For example, the sentence could be re-written "With a telescope, I saw the man." An exemplar approach using statistics derived from a large dataset might be able to suggest a more likely possible interpretation, but since both interpretations are possible it would be impossible to do more than suggest that one interpretation is more likely than another. In my limited exposure to the corpus (limited compared to Google) I would suggest that it's likely the man has a telescope, rather than I used a telescope to see the man. But this relies on my knowledge of telescopes more than it does how many times I've seen sentences like this one. I know that telescopes are used for looking at very distant objects. If the sentence used the word binoculars instead of telescope then it would be a different proposition altogether. I really wouldn't have any way to resolve the ambiguity, and I don't think Google would be able to do much better.

/6

4. The exemplar approach to representing linguistic knowledge uses probability to rank the likelihood of producing or understanding utterances. Describe a situation in which a simple probability calculation would be insufficient to adequately characterize the linguistic situation. (5 points)

An article on translation or linguistics would be very difficult for a probabilistic approach. For example, as Chomsky points out, "Colorless green ideas sleep furiously." is a grammatical sentence that had probably not appeared in print before he wrote it. How can a probability ranking help when the various *n*-grams in the sentence have either not appeared in the corpus at all, or as in the case of "green ideas" have appeared in a specific context that has nothing to do with linguistics? Norvig's talk on translation included examples of sentences being translated into other languages and how Google's model would assign probabilities to the possible alternatives. If the article included unusual translations of sentences how would the probabilistic model know to keep them without understanding that this was an article that included unlikely sentences?

/5

5. Give one example of a system that represents several facts, and that employs representations of these facts that are overlapping, but not holographic. Briefly explain your answer. (6 points)

A web browser history is an example of overlapping, but not holographic representation. For instance, by examining a particular individual's browser history we could learn a lot about the interests of that person, how often and when that person visits certain sites. There would be enormous overlap since certain sites would be visited repeatedly. The pages might be different, but the site the same. For instance, we might learn that the individual visits the NY Times site frequently, but obviously most page views would be unique. We might learn that the most-emailed article list on the NY Times site comes up frequently and this would tell us something about that individual. We might notice many sites covering a particular topic such as cognitive science. We might see a variety of pages on a particular medical condition and be able to surmise something about the individual.

However, any particular piece of information from the browser history would not tell us something about the entire individual and therefore it is not holographic. A single visit to the NY Times homepage on a particular date would not tell us anything about the individual's interest in cognitive science or medical condition.

/6

6. Briefly describe an experimental procedure that might be used to test the psychological reality of the two route model of English past tense formation. (6 points)

The two route model refers to the dual pathways from input stem to output of past tense. One pathway is: input stem → rule → output past tense. The alternate pathway is: input stem → exceptions → output past tense. An experimental procedure to test the psychological reality of this model might involve training children to learn novel verbs, some that are regular in their past tense formation and some that are irregular. First the experimenters would need to establish that the children have a rule for past tense formation that they follow. To do this they could introduce an occasional pseudo verb into a list of real verbs to see if the child will follow a pattern or rule to guess the past tense of the pseudo verb. Then, the experimenters would teach the children the new verbs. If the children consistently apply a past tense formation rule to the new verbs without learning that some are irregular this would be counter evidence to the two route model. If, however, the children learned the irregular verbs along with the regular verbs this might be evidence in favour of the two route model. The experimenters could measure the time taken to form the past tense of a novel verb. Measurement of the difference in time for regular and irregular verbs could provide evidence of a two route model, since presumably it would be more time consuming to consult a "table of exceptions" than to apply a rule.

(ref.: Plunkett & Marchman. From rote learning to system building: acquiring verb morphology in children in connectionist nets. Cognition (1993) p. 21-69)

/6

7. In your own words, briefly describe the distinction between the "two cultures of statistical learning" that Norvig talks about in his essay, "On Chomsky and the two cultures of statistical learning." (6 points)

The two cultures referred to in the article's title are the 'data modeling culture' and the 'algorithmic modeling culture'. Norvig claims that Chomsky's objections (see below) are really objection to the algorithmic modeling culture which aims to develop algorithms that estimate the functions that map inputs to outputs in nature, but that don't necessarily reflect the true underlying nature. In the data modeling culture it is the job of the statistician to choose a model that reflects the reality of nature. The data modeling culture holds that this underlying model is relatively simple.

In the article Norvig quotes Chomsky as saying "There is a notion of success ... which I think is novel in the history of science. It interprets success as approximating unanalyzed data." Norvig, in trying to clarify Chomsky's position, claims that Chomsky believes that the performance of a statistical model is irrelevant if there is no understanding of how the model works beyond statistical means. In other words, if there are no principles that are derived from the model or the analysis of the data then the model's 'success' is irrelevant to science. But Norvig points out that while engineering success is not the goal of science it often goes hand in hand with the advancement of scientific understanding.

Norvig claims Chomsky thinks statistical models are incomprehensible and therefore provide no insight. Norvig counters this claim by admitting that understanding a model with a billion parameters is beyond the comprehension of anyone's brain, but by examining the properties of the model one can gain insight.

/6

8. In last term's Wednesday COGS 300 lab there were 10 males and 8 females. In the Monday COGS 300 lab there were 5 males and 11 females. To balance the numbers, a student was randomly selected from the Wednesday lab and transferred to the Monday lab. Subsequently, a student was randomly selected from the Monday lab and, it turned out, the student selected was male. What now is the probability that the student originally transferred from the Wednesday to Monday lab was male? [Hint: Use Bayes' theorem.] (6 points)

Let H be the hypothesis that the student initially selected from the Wednesday lab was male.
Let E be the evidence that the student subsequently selected from the Monday lab is male.

We can define the following straightforward probabilities:

$$P(H) = \frac{10}{18} \quad P(E \mid H) = \frac{6}{17}$$

A little more work is needed to calculate the last probability required for input into Bayes' formula:

$$P(E) = P(E \mid H) + P(E \mid \overline{H}) = \frac{6}{17} \cdot \frac{10}{18} + \frac{5}{17} \cdot \frac{8}{18} = \frac{60+40}{306} = \frac{100}{306}$$

Thus, Bayes' formula gives us:

$$P(H \mid E) = P(H)\frac{P(E \mid H)}{P(E)} = \frac{10}{18} \cdot \frac{\frac{6}{17}}{\frac{100}{306}} = \frac{3}{5}$$

I find it easier to keep track of the probabilities and see the relationships by constructing a table:

Let m_1 be the event that a male was initially selected from the Wednesday lab.
Let m_2 be the event that a male was subsequently selected from the Monday lab.
Let f_1 be the event that a female was initially selected from the Wednesday lab.
Let f_2 be the event that a female was subsequently selected from the Monday lab.

		E or m_2	$\overline{E}$ or f_2
H	$P(m_1) = \frac{10}{18}$	$P(m_2 \mid m_1) = \frac{10}{18} \cdot \frac{6}{17} = \frac{60}{306}$	$P(f_2 \mid m_1) = \frac{10}{18} \cdot \frac{11}{17} = \frac{110}{306}$
$\overline{H}$	$P(f_1) = \frac{8}{18}$	$P(m_2 \mid f_1) = \frac{8}{18} \cdot \frac{5}{17} = \frac{40}{306}$	$P(f_2 \mid f_1) = \frac{8}{18} \cdot \frac{12}{17} = \frac{96}{306}$
		$P(m_2) = \frac{40}{306} + \frac{60}{306} = \frac{100}{306}$	$P(f_2) = \frac{110}{306} + \frac{96}{306} = \frac{206}{306}$

Now, it's easy to see that $P(m_1 \mid m_2) = \frac{P(m_2 \mid m_1)}{P(m_2)} = \frac{\frac{60}{306}}{\frac{100}{306}} = \frac{60}{100} = \frac{3}{5}$.