9.19 Homework 4
due 1 November 2023

18 October 2023

There's no Colab notebook for this pset: just one `Tregex` tree-search problem, and three problems that you should be able to do with pencil, paper, and in some parts a calculator. Please submit all answers in a single PDF file. Make sure you start all parts of the assignment well before the due date.

Points for each problem: Problems 1 and 2 are worth 35 points each; Problem 3 is worth 20 points (10 for each of the two parts); Problem 4 is worth 10 points.

1 `Tregex` tree searches

For this problem, when we refer to the *Penn Treebank* (Marcus et al., 1994) you can use the data file `wsj-all.mrg` which appears under **Files — Datasets** on Canvas. Also, you can consult the *Penn Treebank Bracketing Guidelines* in the same location to answer questions about how the dataset is organized. The `Tregex` software for conducting tree searches can be downloaded at https://nlp.stanford.edu/software/tregex.shtml. For all parts of the below problem, make sure to **show your work**, including what `Tregex` patterns you used for searches and the math you use for estimating probabilities.

Note: the `Tregex` GUI requires a lot of memory to do searches on big datasets, but the default command invocation gives it "only" 300MB. You can give it more memory by editing `run-tregex-gui.bat` (Windows) or `run-tregex-gui.command` (Mac) and replacing `-mx300m` with a larger value like `-mx2g` (giving it 2 gigabytes). Command-line invocation of `Tregex` generally does not require such a large memory footprint because it streams the output rather than saving the results of searches.

We'll motivate this problem with some examples of possible garden pathing, building off cases discussed in Gibson (1991):

(1) a. Mary's friend gave her books...
    b. Mary's friend gave her books **to the library.**

(2) a. Mary's friend sold her books...
    b. Mary's friend sold her books **to the used book store.**

In the above examples, you may find that you were more surprised by Example (1) the CONTINUATION (1-b) after seeing the PREAMBLE (1-a), than you were by Example (2)'s continuation, (2-b), after seeing the preamble (2-a). This homework assignment investigates why you might have had that experience.

To start the investigation, we need to teach you a bit about the syntax of verb phrases in English. In the DITRANSITIVE construction, there are two different syntactic frames, as illustrated by (nearly) meaning-equivalent alternants like:

Double-Object Construction		**Prepositional-Dative Construction**
Pat gave Kim the book.	vs.	Pat gave the book to Kim.

The **Prepositional-Dative** is so called because the "goal" argument of the verb (the entity that gets given something) is expressed as a prepositional phrase. The **Double-Object** construction is so called because both the goal and the theme (the entity that is being given) are expressed as object NPs, without being marked using a preposition.

1. What is the difference in Penn Treebank syntactic structure between the double-object and prepositional-dative construction in terms of the VP rewrite rule used for each construction? You can answer this by, for example, using the `Tregex` GUI to browse trees containing *give* or other such verbs.

2. There are many different verbs besides *give* that allow both the double-object and prepositional-dative variants, including *sell*, *award*, and *tell*. First, use your intuition to guess the order of preference strength for each of these verbs, ranging from most double-object-preferring to most prepositional-dative-preferring. Report these intuitions. Now, use `Tregex` to compute relative-frequency estimates of

$$P(\text{double-object}|\text{verb with ditransitive meaning})$$

   for various verbs—minimally each of *give*, *sell*, *award*, *tell*, and optionally for any other verbs you want to check—assuming that

$$\begin{aligned} &P(\text{double-object}|\text{verb with ditransitive meaning}) \\ +&P(\text{prepositional-dative}|\text{verb with ditransitive meaning}) \\ =&1 \end{aligned}$$

   —that is, that **double-object** and **prepositional-dative** are the only two ways of realizing a ditransitive meaning. How do the relative frequency estimates line up with your intuitions? Comment on what you've discovered. **Hint:** `Tregex` patterns are left-associative: for example, `@VP < @NP < @NNS` means a VP that has both an NP daughter and an NNS daughter, not a VP that has an NP that has an NNS daughter. You can use parentheses to disambiguate. For example, `@VP < @NP < @NNS` and `(@VP < @NP) < @NNS` mean the same thing, but `@VP < (@NP < @NNS)` means something different (it means a VP that has an NP that has an NNS daughter).

3. Now put together what you've learned about the ditransitive construction with what you've learned about garden-pathing and surprisal to hypothesize why a comprehender might be more surprised by the continuation (1-b) given the preamble (1-a) than by the continuation (2-b) given the preamble (2-a).

2 Surprisal in a simple PCFG of Japanese

Japanese is a strictly verb-final language with null pronouns and considerable word-order freedom. It uses case marking to distinguish subjects (which receive nominative case) from objects (which receive accusative case). Let's consider the following "toy" fragment probabilistic context-free grammar (PCFG) of Japanese. In this fragment, case annotations such as NP[nom] simply distinguish among atomic categories – so NP[nom] and NP[acc] are simply two different categories in the grammar.

0.4	S → NP[nom] NP[acc] V	1.0	NP[nom] → NP *ga*
0.3	S → NP[nom] V	1.0	NP[acc] → NP *o*
0.1	S → NP[acc] NP[nom] V	0.5	NP → *Tanaka-san* ('Mr. Tanaka')
0.2	S → NP[acc] V	0.5	NP → *Ota-san* ('Mr. Ota')
		0.5	V → *deta* ('went out')
		0.5	V → *yonda* ('called')

1. Given the input prefix *Tanaka-san ga*, what is the probability that the next element that the comprehender encounters in the input will be the clause-final verb?

2. Compute the probability that the next element will be the clause-final verb for the input prefixes *Tanaka-san ga Ota-san o*, *Ota-san o*, and *Ota-san o Tanaka-san ga* as well.

3. Let us call the first NP in a sentence NP_1 and the second NP in the sentence, if one appears, NP_2. In each of the nominative-initial and accusative-initial examples, how many **bits of surprisal** that would be associated with encountering the clause-final verb immediately after encountering NP_1 are reduced by encountering NP_2 first? Remember that surprisal is negative log probability, or log inverse-probability:

$$\text{Surprisal of } x \text{ in context } C = -\log_2 P(x|C)$$
$$= \log_2 \frac{1}{P(x|C)}$$

so the bits of surprisal reduced by NP_2 would be

$$\log_2 \frac{1}{P(x|\text{NP}_1)} - \log_2 \frac{1}{P(x|\text{NP}_1\ \text{NP}_2)}$$

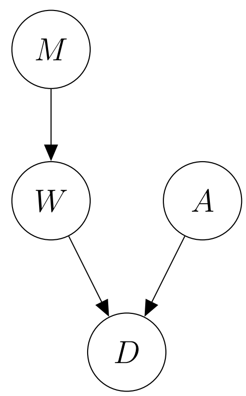
4. Could you imagine a reasonable PCFG (for any language) in which adding a pre-verbal constituent would ever *increase* the amount of surprisal associated with the final verb? Justify your answer.

3 Conditional Independence in Bayes Nets

For each Bayes Net below, we give you several pairs of sets of nodes in the Bayes Net. In each case, state the conditions (what sets of nodes must and/or must not be **known**) under which the specified node sets will be conditionally independent from one another. If the node sets are always independent or can never be independent, say so.

1. A variant of the disfluency model we saw earlier:

M	intended meaning to be conveyed
W	is the word intended to be spoken a hard word?
A	was the speaker's attention distracted?
D	was a disfluency uttered?

   (a) $\{W\}$ and $\{A\}$

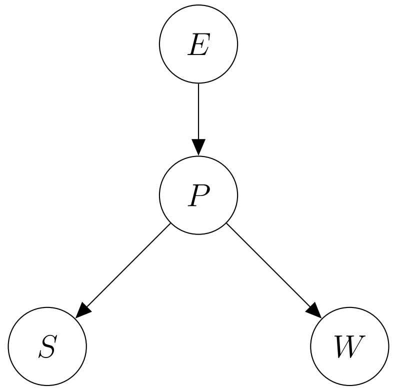
   (b) $\{M\}$ and $\{D\}$

   (c) $\{M\}$ and $\{A\}$

   (d) $\{D\}$ and $\{A\}$

2. The relationship between a child's linguistic environment, his/her true linguistic abilities/proficiency, and measures of his/her proficiency in separate spoken and written tests

E	a child's linguistic environment
P	the child's linguistic proficiency (number of words known, etc.)
S	the child's performance on a spoken language proficiency test
W	the child's performance on a written language proficiency test



(a) $\{S\}$ and $\{W\}$

(b) $\{E\}$ and $\{P\}$

(c) $\{E\}$ and $\{S\}$

(d) $\{E, P\}$ and $\{S\}$

(e) $\{E,P\}$ and $\{S,W\}$

4 Conditional Independence in the Bayesian perceptual magnet model

In the perceptual magnet example from the Bayes Net class, did any CONDITIONAL INDEPENDENCE hold among the three random variables of the model—category c, target production T, and perceived speech signal S? (**Hint:** you may want to review the characterization of the Bayesian perceptual magnet model as a noisy-channel model on the perceptual magnet slides we covered in class, and/or review the mathematical presentation of the model on page 758 of Feldman et al. (2009).)

References

Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*(4), 752–782.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown* (Doctoral dissertation). Carnegie Mellon.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1994). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*(2), 313–330.