9.19/9.190: Computational Psycholinguistics, Pset 1
due 25 September 2023

8 September 2023 (updated 18 September 2023)

Incremental inference about possessor animacy

English has two CONSTRUCTIONS for grammatically expressing possession within a noun phrase, as exemplified in (1)–(2) below:

(1) the queen's crown (PRENOMINAL or 'S GENITIVE: possessor comes before the possessed noun)

(2) the crown of the queen (POSTNOMINAL or *of* GENITIVE: possessor comes after the possessed noun)

There is a correlation between the ANIMACY of the possessor and the preferred construction: animate possessors, as above, tend to be preferred prenominally relative to inanimate possessors, as in (3)– (4) below (Futrell & Levy, 2019; Rosenbach, 2005):

(3) the book's cover (Prenominal)

(4) the cover of the book (Postnominal)

Here is a pair of conditional probabilities that reflects this correlation:

$$P(\text{Possessor is } \textbf{prenominal} | \text{Possessor is } \textbf{animate}) = 0.9$$
$$P(\text{Possessor is } \textbf{prenominal} | \text{Possessor is } \textbf{inanimate}) = 0.25$$

Now consider the cognitive state of language comprehenders mid-sentence who have heard each of the three respective example sentence fragments, where the nouns that have been uttered are unfamiliar words:

1) the sneg of...
2) a...
3) a tufa's dax...

Task: Based on the knowledge encoded in the probabilities above, plot the probability in each of these three cases that the comprehender should assign to the possessor being animate, as a function of the prior probability $P(\text{Possessor is } \textbf{animate})$. That is, your plots will have

the prior probability P(Possessor is **animate**) on the x-axis, and the posterior probability P(Possessor is **animate**|the provided sentence fragment) on the y-axis. Show your work in setting up the computations.

To get you started, we have created a simple Colab notebook that shows you how to create plots, here:

https://colab.research.google.com/drive/1yxQC2Hg52gDUvoKfaTEgX6dhbEsvTycM?usp=sharing

If you want to use a different programming language than Python to generate the plots, however, that is fine.

Phoneme categorization

The questions in this section relate to ideal probabilistic categorization of instances of the sound categories /b/ and /p/, as covered in Lecture 1 (related readings include Clayards et al., 2008, Feldman et al., 2009).

Assume that a single informative cue (VOT) distinguishes between these categories, and that the distributions of VOT values for these categories can be approximated by Gaussian distributions with means of $\mu_b = 0$ and $\mu_p = 50$. Imagine a context in which the prior probabilities of the two categories differ, $p(/\text{b}/) = 0.75$ and $p(/\text{p}/) = 0.25$.

For a given VOT value x, we can calculate the posterior distribution on the category c that token came from $p(c|x)$ using Bayes rule:

$$p(c|x) = \frac{p(x|c)p(c)}{p(x)} \tag{1}$$

$$= \frac{p(x|c)p(c)}{\Sigma_{c'} p(x|c')p(c')} \tag{2}$$

where the prior $p(c)$ is as given above, the likelihood $p(x|c)$ is given by the Gaussian probability density function

$$p(x|c) = \frac{1}{\sigma_c \sqrt{2\pi}} \exp\left[-\frac{(x-\mu_c)^2}{2\sigma_c^2}\right] \tag{3}$$

and the normalizing constant in the denominator is evaluated by summing across all possible hypotheses $c' \in \{/\text{b}/, /\text{p}/\}$:

$$p(c|x) = \frac{p(x|c)p(c)}{p(x|/\text{b}/)p(/\text{b}/) + p(x|/\text{p}/)p(/\text{p}/)} \tag{4}$$

1. Imagine that both categories had equal variances $\sigma_b^2 = \sigma_p^2 = 144$. Under this assumption, the posterior probability of the category /p/ for a VOT value of 25 ms, i.e., $p(c = /\text{p}/|x = 25\text{ms})$, is easy to calculate. Why is the posterior probability easy to calculate, and what is it? Now, plot the posterior for VOT values ranging from $-25ms$ to 75 ms.

2. In fact, VOTs for voiceless stops such as /p/ are more variable than those for voiced stops such as /b/. This means that the Gaussian approximations of these categories should have different variances, such as $\sigma_b^2 = 64$ and $\sigma_p^2 = 144$. Assuming these values, plot the posterior for the range you used in part (1). For a VOT value of 25 ms, how has the categorization preference changed, and why?

3. Continuing to assume the unequal-variance parameters as in (2), guess the posterior for the very low VOT of −200ms, and then calculate it. There is some counter-intuitive behavior: what is it? What does this counter-intuitive behavior tell us about the limitations of the model we've been using? *Optionally, extend your continuous plots down to a VOT of* −200 *ms to see how this counter-intuitive effect develops.*

Testing a state-of-the-art speech recognition model with phonetically ambiguous sentences or phrases

During the September 13 class session, we spent some time coming up with phonetically ambiguous sentences—that is, sentences that sound close enough to another sentence that they are easily mistaken for a phonetically similar or identical sentence. Examples include *It's not easy to wreck a nice beach* (confusable with *It's not easy to recognize speech*) and *Phil and Mary are our young cousins* (confusable with *Phil and Mary are young cousins*). We also ran a couple of these examples through OpenAI's Whisper speech recognition model. For this problem, your task is to do a version of this at home:

1. Come up with a new example of a phonetically ambiguous sentence or phrase;

2. Explain the nature of the phonetic ambiguity;

3. Record yourself speaking both the sentence and its "neighbor" sentence with which it is potentially confusable;

4. Run these recordings through Whisper to see how the model transcribes them.

Two important "resources" for you in constructing your example are (a) homophones (words that sound the same but mean different things, like *piece* and *peace*) and (b) differing locations of word boundaries, such as *I scream* versus *ice cream*. Ideally your example will involve more phonetic ambiguity than a single case of homophony. Whisper has five model sizes—`tiny, base, small, medium, large`. Please report the transcriptions made by all five models. (The `large` model has only one version: multi-lingual. The other model sizes have english-only and multilingual both. You can choose which version to use.)

Also, the phonetic ambiguity of a sentence or phrase may depend on whether it is spoken casually or more carefully, as there can be a lot of phonetic reduction in casual speech. If your example requires casual speech to be phonetically ambiguous, then do two versions of the recordings of both sentences: one version spoken casually, and one version spoken more carefully, and report Whisper's transcriptions of all of them.

We have prepared a Google Colab notebook for running Whisper, which you can access here:

https://colab.research.google.com/drive/1sHVtIeX1JRahpiLrW0o0aW37zkoorpNW?usp=sharing

For your submission be sure to include the recorded audio file(s). Also, if you want to do this problem in a language other than English, you are more than welcome to do so!

References

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*, 804–809.

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.

Futrell, R., & Levy, R. P. (2019). Do RNNs learn human-like abstract word order preferences? In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*.

Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *Language*, *81*(3), 613–644.