9.19/9.190: Computational Psycholinguistics, Pset 1
due 22 September 2023

8 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 comprehender mid-sentence who have heard each of the three respective example sentence fragments, where the nouns that have been uttered are unfamiliar words to the comprehender:

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(Possessor is **animate**). Show your work in setting up the computations.

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, $p(/\mathrm{b}/) = 0.75$ and $p(/\mathrm{p}/) = 0.25$.

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 $c' \in \{/\mathrm{b}/, /\mathrm{p}/\}$:

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

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., $p(c = /\mathrm{p}/|x = 25\mathrm{ms})$, 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.*

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.