096222: Language, Computation and Cognition
Homework Assignment 1
due 20 April 2025

31 March 2025

1 Your lifetime linguistic exposure

Estimate how many Hebrew words you have *heard* and *read* in your lifetime to date. Your calculation should be for the number of *tokens*, not *types* – i.e., every exposure to a word counts as an additional token even if it is a repetition of a familiar word type. Provide a detailed justification for your answer. Your reasoning is most important to see, not the exact number that you come up with.

In addition to the written explanation, fill in your estimate for the total number of words you were exposed to (heared + read) in the Words, words and more words Moodle form. We will present the class-level statistics during the lecture.

2 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 set of probabilities that reflects this correlation:

$$P(\text{Possessor is } \textbf{animate}) = 0.4$$
$$P(\text{Possessor is } \textbf{prenominal} | \text{Possessor is } \textbf{animate}) = 0.9$$
$$P(\text{Possessor is } \textbf{prenominal} | \text{Possessor is } \textbf{inanimate}) = 0.2$$

Now consider the cognitive state of a language comprehender mid-sentence who has just heard the start of a noun phrase involving a noun that they don't know:

the sneg of the...

Task: Based on the knowledge encoded in the probabilities above, what probability should the comprehender assign to the upcoming possessor being animate? Show your work in carrying out this computation.

3 Entropy

Imagine a child language learner in a very, very early stage of acquisition with a four-word vocabulary following the below word frequency distribution:

Word	Unigram probability
w_1	$\frac{1}{2}$
w_2	$\frac{1}{4}$
w_3	$\frac{1}{8}$
w_4	$\frac{1}{8}$

What is the **entropy** of the unigram distribution for this vocabulary?

What is the **maximum possible unigram entropy** for a four-word vocabulary? Explain your answer.

4 The N400 ERP component

You are transported back to a cognitive science lab in 1979.[1] A researcher in the lab explains that the lab wants to characterize the N400 response. The researcher continues, "Our hypothesis is that semantic incongruities cause it. The experimental design that I had in mind is to compare normal sentences with sentences that have an unexpected word. Example sentences are 'He took a sip from his drink' and 'He took a sip from the transmitter.' If the unexpected words consistently result in a larger N400 response than the expected ones, then we will have proven the hypothesis to be correct. What do you think?"

Why might the above experimental design not necessarily prove the hypothesis? Name another factor that could also result in the expected outcome. Describe what additional sentences you could compare or how else you could control for this factor to better prove the hypothesis.

[1] Fun fact: The first paper on the N400 was published in January, 1980: Kutas and Hillyard (1980).

5 Visual World Paradigm

Read Tanenhaus et al. (1995) and provide at least one question or comment on the paper using Perusall. Your contribution(s) can be part of a discussion initiated by other students.

References

Futrell, R., & Levy, R. P. (2019). Do RNNs learn human-like abstract word order preferences? *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, *2*, 50–59.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

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

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.