Bias in the cloze task

Nathaniel J. Smith <njsmith@cogsci.ucsd.edu>
UC San Diego

Roger Levy <rlevy@ling.ucsd.edu>
UC San Diego

What does the cloze task actually measure?

Studies often use the cloze task (Taylor, 1953) to estimate the predictability of a continuation given some context. This takes advantage of participants' rich knowledge of their language, but it also assumes that participants' probability match, that is, if they produce responses by sampling from their subjective distribution over potential continuations.

Is this assumption justified, or do the peculiarities of the cloze task produce distortions? We compare experimentally cloze responses to "gold standard" corpus-based probability distributions, and find that many psycholinguistic properties bias cloze responses. This produces the potential for spurious results in studies which use cloze to estimate probability.

Methods

Stimuli

4-word sentence initial stems from the Google Web 1-T corpus of 5-grams, with enough occurrences in the corpus that no smoothing assumptions were needed. Mostly (99%) open-class continuations. A wide range of continuation structures (probability of the most likely and second-most likely continuations were parametrically varied).

Careful screening, both automated and by hand, was necessary to remove "Web lingo"; the most common sentence in English is "All rights reserved." Phrases used predominantly in technical genres; in another series of Spera, continuations which happen to occur in sampled text are over-counted. Phrases which were "too similar"; e.g., where only pronoun gender differed.

Total stems: 300
Each occurred a minimum of 250 times, median of 1966 times.

Participants

Native English speakers from UCSD subject pool.
128 participants performed the sentence completion task online.
26 performed the same task in a lab environment.

The results are the same, so we collapse them (total N = 134).

Analysis

We examine two types of effects:
An overall shift in probability mass, the degree of which may be affected by stem:

- Bias towards or away from individual words based on their psycholinguistic properties:

- Cloze is biased toward familiar, concrete words, with high contextual diversity, that are infrequent/acquired late

Discussion

We find that cloze probabilities differ greatly from web corpus probabilities, and that these differences correlate with common psycholinguistic variables.

Some results here may be artifacts of using web-derived stimuli, but suppose some are correct. Then:

- If participants in a reading study are sensitive to true probability... but we control for cloze, rather than true probability... then the above variables will "correct" our cloze estimates and appear to matter...
- even if readers don't care about them at all.

(Or, if readers are sensitive to cloze, then corpus probabilities have this problem.)

Next step: use these stimuli in a self-paced reading study, to check whether corpus or cloze probability better predicts reading time.

CUNY 2010: New York
March 19, 2010