2015 LSA Summer Institute: Computational Psycholinguistics

Homework #3

Klinton Bicknell (Northwestern) and Roger Levy (UCSD)

Tuesday 14 July 2015
Due: Friday 17 July 2015

Let us review the logic of Uniform Information Density (Levy & Jaeger, 2007; Jaeger, 2010) as applied to non-subject-extracted relative clauses (NSRCs) in English, as

(1) I know the family (that) you cook for.

In the relativizer-free variant, several pieces of information are conveyed by the onset of the NSRC—the word *you*:

1. A relative clause has begun, with probability ≈ 1 (try continuing the sentence beyond *you* without constructing a relative clause!);
2. the relative clause is not subject-extracted;
3. the subject of the relative clause is *you*.

If we use information theory to quantify these three pieces of information via log-probabilities, it follows the chain rule decomposition and is truly additive:

$$\begin{aligned} P(you|\text{Context}) = & P(\text{RC subject is } you|\text{not SRC}, \text{RC}, \text{Context}) \times & (1)\\ & P(\text{not SRC}|\text{RC}, \text{Context}) \times & (2)\\ & P(\text{RC}|\text{Context}) & (3)\\ \log P(you|\text{Context}) = & \log P(\text{RC subject is } you|\text{not SRC}, \text{RC}, \text{Context}) + & (1)\\ & \log P(\text{not SRC}|\text{RC}, \text{Context}) + & (2)\\ & \log P(\text{RC}|\text{Context}) & (3) \end{aligned}$$

In the version with a relativizer, this information is split up: piece (1) lives on the relativizer, and (2) and (3) live on the first word of the RC. Therefore inserting the relativizer strictly reduces the information density of the RC onset. The UID prediction follows that relativizers should be used most often at informationally dense RC onsets.

Question: Let us now consider the case of a verb-final language with word order variation, such that orders SOV and OSV are both possible. A comprehender will encounter the

subject and object before the verb. In general, the identity of subject and object will not be conditionally independent given the context preceding the utterance. The joint distribution of subject and object given context is $P(\mathrm{S}, \mathrm{O}|\mathrm{Context})$. What UID-based predictions can you derive regarding when SOV versus OSV word order will be preferred? You can use any combination of intuitive and/or mathematical reasoning that you like in answering this question.

(**Hint 1:** recall that the chain rule decomposition can be applied to the joint distribution in either order:

$$P(\mathrm{S}, \mathrm{O}|\mathrm{Context}) = P(\mathrm{S}|\mathrm{O}, \mathrm{Context})P(\mathrm{O}|\mathrm{Context})$$

or

$$P(\mathrm{S}, \mathrm{O}|\mathrm{Context}) = P(\mathrm{O}|\mathrm{S}, \mathrm{Context})P(\mathrm{S}|\mathrm{Context})$$

.)

(**Hint 2:** very generally, the log of an inverse is the negative of the log:

$$\log \frac{1}{x} = -\log x$$

and taking a logarithm transforms products into sums:

$$\log [p \times q] = \log p + \log q.$$

This means, for example, that logging a chain rule decomposition leads to a sum of log-probabilities—or equivalently of surprisals:

$$\begin{aligned}
P(X, Y) &= P(Y|X)P(X) && \text{(chain rule decomposition)} \\
\log P(X, Y) &= \log P(Y|X) + \log P(X) && \text{(logs transform products into sums)} \\
-\log P(X, Y) &= -\log P(Y|X) - \log P(X) \\
\log \frac{1}{P(X, Y)} &= \log \frac{1}{P(Y|X)} + \log \frac{1}{P(X)} && \text{(negative of a log is log of the inverse)}
\end{aligned}$$

.)

References

Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology*, *61*, 23–62.

Levy, R. & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In *Proceedings of the 20th conference on Neural Information Processing Systems (NIPS)*.