

Name:

Statistical Parsing & Surprisal

Tutorial 8

The Parser Files

- The files and programs you will need are in:
/proj/courses/comppsy/Tutorial8/
Copy the entire directory to your account.

For this tutorial we will again experiment with the top-down incremental probabilistic parser developed by Brian Roark. While not explicitly developed as a cognitive model, it allows us to examine the kinds of behavior the parser exhibits for some of the lexical, structural, and combined ambiguities we've discussed in the lecture.

1. Surprisal

The `-p` switch tells the parser to output more detailed information on a word-by-word basis, including prefix probabilities, surprisal, as well as lexical and syntactic surprisal (which sum to give total surprisal). For this question, test the reduced and unreduced relative clause examples discussed in the lecture:

```
% ./tdp -v -p -F rels.out /proj/contrib/tdp.distrib/parse.model.slc.p05  
rels.txt
```

Use the `stanford-tregex-2012-07-09` (from last tutorial) tool to confirm that the parser got the correct parse for both sentences. Does surprisal correctly explain the garden-path effect on “fell”, and Hale demonstrated.

2. Subject/Object relative clauses

Now do the same to determine whether surprisal (lexical and/or syntactic) explains the relative ease of processing subject over object relative clauses, using the sentences in `subjobj.txt`

3. NP/S Complement Ambiguity

The file `nps.txt` contains examples of the NP/S ambiguity discussed in class. Is there any evidence in surprisal scores that the parser uses the implausibility of “shoes” to disambiguate towards the S complement reading? You might want to try looking at the incremental parsers too. Summarize the behavior of the parser for these examples.

4. Average parser performance

Construct three additional sentence pairs similar to the (reduced) relative sentences in Q2, but with different lexical items. Determine the parsers' measure for lexical, syntactic and total surprisal, and average these for both sentence types across all 4 versions.

Plot the average of all three surprisal values, for the reduced and unreduced versions, at the first verb (e.g. *raced*), and the second verb (e.g. *fell*). Does the average reveal similar surprisal effects as you observed in Q1?

5. Information Density

Recall from the lecture the claim that “that” can be omitted when the onset of the complement clause is less surprising, as in the case when following a verb that usually takes an S complement.

Examine the surprisal measures of the parser for the sentence in `thatcomp.txt` – to what extent is the UID hypothesis confirmed or disconfirmed?