Expectation-based syntactic processing

- Structures we expect are easy to process
- Modern computational linguistics techniques → precise psycholinguistic model
- Matches empirical results of four experiments better than traditional memory-based models
- A new information-theoretical derivation of an elegant model (Hale 2001)

Contrasted against two memory-based syntactic processing models
1. Syntactic Prediction Locality Theory (SPLT; Gibson 1998): more & more distant dependencies are hard

2. Active Filler Hypothesis (AFH: Clifton & Frazier 1989): fillers cause a greedy search for a gap

Hale 2001: a distinctive ranked-parallel model
- The more a word is expected, the easier it is to process: difficulty ~ SURPRISAL(w)
- SURPRISAL(w) = -log P(w|CONTEXT)
- Parallel parsing with a probabilistic context-free grammar (PCFG) determines the word expectation P(w|CONTEXT)
- Algorithms (Lafferty & Jelinek 1992; Stolcke 1995) give us p(w) from a PCFG
- Hale’s surprisal theory can also be derived on information-theoretical grounds (see box below)

Derivation of surprisal from relative entropy:
- sentence comprehension: choosing the best [most probable] syntactic/semantic structure from among possible structures (T)
- Partial input w induces a preference distribution [probability distribution] D over possible T
- D must be constantly updated (for inference!)
- Suppose greater changes in D incur greater cost
- Operationalize as the relative entropy between distributions BEFORE and AFTER w
- Relative entropy over trees comes out as surprisal over words!
- Connects disambiguation with processing difficulty
- Representation-independent: processing model now consists solely of a model p(w)

Clause-final verbs: reading time

- Memory-based prediction: more preverbal dependents increase difficulty at final verb
- But several experiments show that these extra dependents decrease reading time
  - Konieczny 2000; Konieczny & Döring 2003; Vasishth 2002
  - Konieczny & Döring 2003: extra dative NP
  - SPLT: final verb read faster in genitive DES condition
  - Observed: final verb read faster in dative DEM condition

incur greater cost

Clause-initial NPs: the subject preference

- German main-clause NP order is free
- So case-syncretized clause-initial NP can be subject or object
- Hemforth 1993: default preference as subject
- Schleswesky et al. 2000: explanation via AFH

Memory-based prediction: more preverbal dependents

AFH predicts difficulty at postverbal NP
- But it actually shows up at the following PP

AFH-predicted locus of difficulty

And PP modification is more likely for object NPs:
- All NPs
- Post-verbal NPs

This results in a prediction that differential difficulty will appear at the PP postmodifier
- This is exactly where it does appear