German word order and expectation-based syntactic processing

Roger Levy, Stanford University

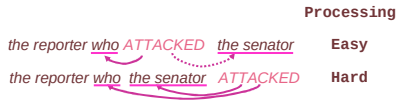
CUNY Sentence Processing Conference, April 1, 2005, Tucson, AZ

Expectation-base syntactic processing

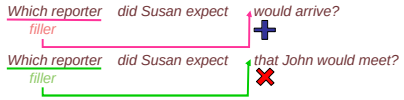
- Structures we *expect* are easy to *process*
- Modern computational linguistics techniques → precise psycholinguistic model
- Matches empirical results of four experiments better than traditional memory-based models
- A new information-theoretical derivation of an elegant model (Hale 2001)

- Contrasted against two ***memory-based*** syntactic processing models

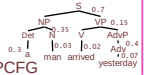
1. Syntactic Prediction Locality Theory (SPLT; Gibson 1998): more & more distant dependencies are hard



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

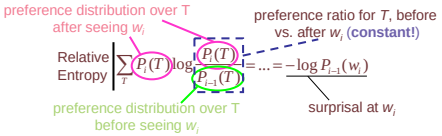


- Hale 2001: a distinctive ranked-parallel model
  - The more a word is expected, the easier it is to process: ***difficulty ~ SURPRISAL(w_i)***
    - **SURPRISAL(w_i) ≡ -log P(w_i\|CONTEXT)**
  - Parallel parsing with a ***probabilistic context-free grammar*** (**PCFG**) determines the word expectation P(w_i\|CONTEXT)
  - Algorithms (Lafferty & Jelinek 1992, Stolcke 1995) give us $p_i(w)$ 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 $w_{1...i}$ 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_i



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 $p_i(w)$

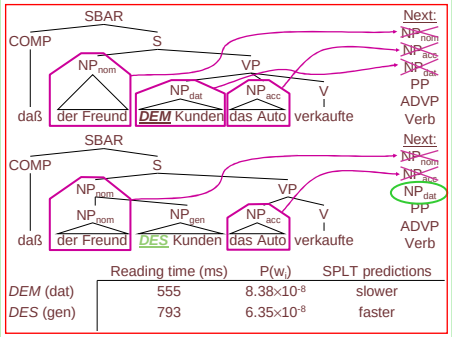
Clause-final verbs: reading time

- 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

*...der Freund **DEM** Kunden das Auto verkaufte*
...the friend the client the car sold
'...the friend sold the client a car...'

*...der Freund **DES** Kunden das Auto verkaufte*
...the friend the client the car sold
'...the friend of the client sold a car...'

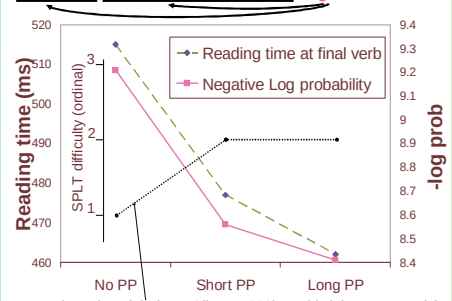
- Expectation-based processing directly predicts faster final-verb reading for more preverbal dependents
  - Having seen more = having more information
  - More information = more tightly constrained expectations



	Reading time (ms)	P(w_i)	SPLT predictions
DEM (dat)	555	8.38×10^{-8}	slower
DES (gen)	793	6.35×10^{-8}	faster

- Surprisal model also successfully predicts effect of extra preverbal verb-modifying PP (Konieczny 2000):

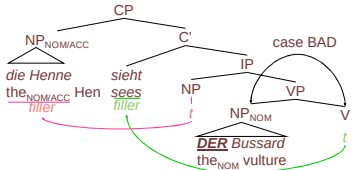
	SPLT Prediction	Result
He has the group led / *Er hat die Gruppe* **geführt**	easy	slow
He has the group to the mountain led / *Er hat die Gruppe auf den Berg* **geführt**	hard	fast
...the group to the very beautiful mountain led / *...die Gruppe auf den sehr schönen Berg* **geführt**	hard	fastest



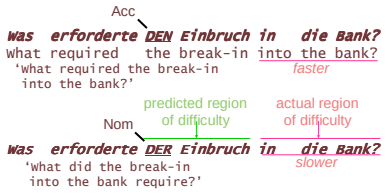
Memory-based models (e.g., Gibson 1998) would violate monotonicity

Clause-initial NPs: the subject preference

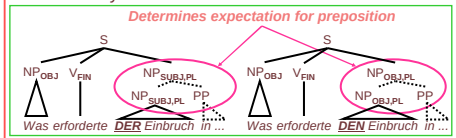
- German main-clause NP order is free
- So case-syncretized clause-initial NP can be subject or object
  - Hemforth 1993: default preference as subject
  - Schlesewsky et al. 2000: explanation via AFH



- Schlesewsky et al. 2000's experiment
  - Clause-initial *was* 'what' is just as often a subject as it is an object
  - postverbal NP article disambiguates
- AFH predicts difficulty at postverbal NP
  - But it actually shows up at the following PP



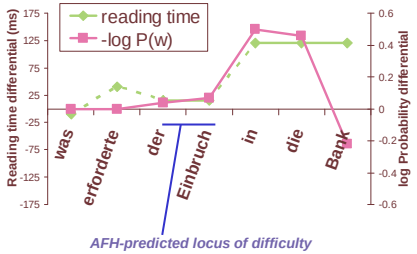
- In incremental parse, expectation for a preposition differs only in the PP modification rule used



- And PP modification is more likely for object NPs:

	All NPs		Post-verbal NPs	
	Subject	Object	Subject	Object
NEGRA	15.3%	22.4%	12.2%	20.3%
TIGER	15.2%	23.7%	12.2%	24.4%
TüBa-D/Z	6.4%	11.3%	4.4%	9.4%

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



rog@stanford.edu http://www.stanford.edu/~rog