Empirical: lexical access, word category/sense, subcategorization

Rational: accurate, robust, broad coverage

Rational Models:
• explain accurate performance in general: i.e. rational behaviour
• explain specific observed human behavior: e.g. for specific phenomena

\[ \text{argmax } P(s_i) \text{ for all } s_i \in S \]

Lexical Category Disambiguation

- Sentence processing involves the resolution of lexical, syntactic, and semantic ambiguity.
- Solution 1: These are not distinct problems
- Solution 2: Modularity, divide and conquer
- Category ambiguity:
  - Time flies like an arrow.
- Extent of ambiguity:
  - 10.9% (types) 65.8% (tokens) (Brown Corpus)

The Model: A Simple POS Tagger

- Find the best category path \((t_1 \ldots t_n)\) for an input sequence of words \((w_1 \ldots w_n)\):
  \[ P(t_0, \ldots t_n, w_0, \ldots w_n) \]
- Initially preferred category depends on two parameters:
  - Lexical bias: \(P(w_i | t_i)\)
  - Category context: \(P(t_i | t_{i-1})\)
- Categories are assigned incrementally: Best path may require revision

![Viterbi explained!](image-url)
2 Predictions

- The Statistical Hypothesis:
  - Lexical word-category frequencies, $P(w_i|t_i)$, are used for initial category resolution

- The Modularity Hypothesis:
  - Initial category disambiguation is modular, and not determined by (e.g. syntactic) context beyond $P(t_i|t_{i-1})$.

- Two experiments investigate
  - The use word-category statistics
  - Autonomy from syntactic context

Statistical Lexical Category Disambiguation

- Initially preferred category depends on: $P(t_0,\ldots,t_n, w_0,\ldots,w_n) \approx \prod_{i=1}^{n} P(w_i|t_i)P(t_i|t_{i-1})$

- Categories are assigned incrementally
  - the warehouse *prices* the beer very modestly
  - the warehouse *prices* are cheaper than the rest
  - the warehouse *makes* the beer very carefully
  - the warehouse *makes* are cheaper than the rest

- Interaction between bias and disambiguation
- Category frequency determines initial decisions
Modular Disambiguation?

- Do initial decisions reflect integrated use of both lexical and syntactic constraints/biases or just (modular) lexical category biases?
  - N/V bias with immediate/late syntactic disambiguation as noun

- Main effect of bias at disambiguation:
  - Initial decisions ignore syntactic context.
  - Problematic for lexicalist syntactic theories
  - At c2, VA/VU difference is significant
  - Implies lexical category doesn’t include number (?)

\[ \text{a) [V-bias, N-disamb]} \quad \text{The warehouse makes are cheaper than the rest.} \]
\[ \text{b) [V-bias, N-unamb]} \quad \text{The warehouse make is cheaper than the rest.} \]
\[ \text{c) [N-bias, N-disamb]} \quad \text{The warehouse prices are cheaper than the rest.} \]
\[ \text{d) [N-bias, N-unamb]} \quad \text{The warehouse price is cheaper than the rest.} \]

‘That’ Ambiguity (Juliano & Tanenhaus)

A. That experienced diplomat(s) would be very helpful ... [DET]
B. The lawyer insisted that experienced diplomat(s) would be very helpful [Comp]

- Corpus based estimates:
  - Initially: det=.35, comp=.11  Post-verbally: comp=.93, det=.06

- Found increased RT when dispreferred POS for “that” (according to context) is forced in the disambiguation region “diplomat(s)”

- Advocates bigram over unigram: \( P(t_i|t_{i-1}) \)

\[
\begin{align*}
P(\text{that}|\text{comp}) &= 1, P(\text{that}|\text{det}) = 0.171 \\
P(\text{comp}|\text{verb}) &= 0.0234, P(\text{det}|\text{verb}) = 0.0296 \\
P(\text{comp}|\text{start}) &= 0.0003, P(\text{det}|\text{start}) = 0.0652
\end{align*}
\]
Internal Reanalysis

• The tagger model predicts internal reanalysis for some sequences.

• Viterbi: revise most likely category sequence based on next transition

• Right context in RR/MV ambiguities: [MacDonald 1994]
  • The sleek greyhound *raced at the track* won the event
  • The sleek greyhound *admired at the track* won the event

• *raced* = intrans bias, *admired* = trans bias

• Increased RT *(blue)* indicate transitivity bias is used

An SLCM Account

• Assume transitive/intransitive POS categories, extract frequencies from the Susanne corpus:

  *The man fought at the police station fainted*  [intransitive]
  *The man held at the police station fainted*  [transitive]
Reduced Relative Clause

• Parsers can make wrong decisions that lead them up the garden path

“The man raced to the station was innocent”


The Problem

• In some cases it may be possible to recover from the error earlier

“The man held at the station was innocent”

A Puzzle

- Sometimes local dependencies appears to violate the global parse:
  - [A/R] The coach smiled at the player tossed a frisbee by the ...
  - [U/R] The coach smiled at the player thrown a frisbee by the ...
  - [A/U] The coach smiled at the player who was tossed a frisbee by the ...
  - [U/U] The coach smiled at the player who was thrown a frisbee by the ...

- Syntactically, “tossed” is must a past-participle in this context, but what do people do?

- We might expect to see:
  - Main effect of verb ambiguity: if ambiguous verbs are difficult
  - Main effect of structure ambiguity: if ambiguous RRCs are difficult

Results:

- These results are problematic for theories requiring global syntactic wellformedness (e.g. Frazier, 1987; Gibson, 1991, 1998)
An SLCM Account

• Initially preferred category depends on two parameters:
  • Lexical bias: \( P(w_i | t_i) \) Category context: \( P(t_i | t_{i-1}) \)

• **[AS-AV]** The coach smiled at the player tossed a frisbee … \([slowest]\]
  
  \[ P(tossed|Vpast) * P(Vpast|noun) > P(tossed|Vpart) * P(Vpart|noun) \]

  • So: assign \( tossed=Vpast \), but can’t integrate into parse, so reanalyse

• **[US-AV]** The coach smiled at the player who was tossed a frisbee … \([fast]\]
  
  \[ P(tossed|Vpast) * P(Vpast|Aux) < P(tossed|Vpart) * P(Vpart|Aux) \]

  • So: assign \( tossed=Vpart \), integrate into parse, no difficulty

---

**SLCM Summary**

• Psychologically plausible: lower statistical complexity than other models

• High accuracy in general: explains why people perform well overall

• Explains where people have difficulty
  • Statistical: category frequency \textbf{drives} initial category decisions
  • Modular: syntax structure \textbf{doesn’t determine} initial category decisions
  • Bigram evidence: “that” ambiguity [Juliano and Tanenhaus]
  • Reanalysis of verb transitivity for ‘reduced relatives’ [MacDonald]
Comments on the SLCM

• Evidence category preference appears truly frequency-based

• Indication of which features are exploited [e.g. transitivity, not number]
  • But this is subject to further empirical investigation & verification

• Combines optimality of probabilities with advantages of modularity
  • Psychological plausibility due to tractable parameter space

• Implications for the Grain Problem?
  • Bigrams used, but not tri-grams, or syntactic structure?
  • Transitivity but not number? More/less syntactically-rich POS tags?

Probabilistic Syntax

• The SLCM is only a model of lexical category assignment
  • But note: these category decisions underlie many “syntactic” ambiguities

• Some ambiguities are purely syntactic, however:
  • Relative clause attachment, or other modifier attachment
  • NP/S complement ambiguity (unless subcat is encoded in the POS tags)
  • Also evidence that compositional interpretation influences parsing
  • Can’t be modeled in the SLCM alone

• Apply probabilistic approaches to modeling human syntactic parsing