So far: All about Syntax

- What algorithms are used to construct syntactic analyses from the grammar?
- What mechanisms are used to deal with lexical & syntactic ambiguity?
  - is it one mechanisms, or are these levels distinct?
  - serial+backtracking, parallel, monotonic?
- What kind of information is used to decide upon/rank alternative analyses?
  - Structural simplicity (Frazier), thematic dependencies (Pritchett), probabilities (Jurafsky)?
- What is the linking mechanisms from the parser to reading measures?
  - Re-parsing cost, revising role assignments, non-monotonicity, parse pruning and re-ranking
- What other factors determine interpretation preferences and processing cost?
Kinds of constraints

• The doctor told the woman that … story
diet was unhealthy
he was in love with her husband
he was in love with to leave
story was was about to leave

• Prosody: intonation can assist disambiguation

• Lexical preference: that = {Comp, Det, RelPro}

• Subcat: told = [ [ _ NP NP] [ _ NP S] [ _ NP S’] [ _ NP Inf] ]

• Semantics: Referential context, plausibility
  • Reference may determine “argument attach” over “modifier attach”
  • Plausibility of story versus diet as indirect object

The Modularity Argument

• There is considerable evidence that non-syntactic information and context influence reading times.

• However there is limited evidence that the core syntactic preferences are ever completely overridden, e.g. (Rayner et al, 1983):
  • The florist sent the flowers was very pleased.
  • The performer sent the flowers was very pleased.

• While there is less of a garden path effect at “was very pleased” for the “performer” condition, it was still greater than for unambiguous controls.

• Thus: “initial” preferences for syntactic analyses are driven by modular, syntactic strategies, with other knowledge being used only “later” s.
Constraint-based Models

- What architecture is assumed?
  - Non-modular: all levels are constructed and interact simultaneously

- What mechanisms is used to construct interpretations?
  - Parallel & competitive: ranking based on constraint activations

- What information is used to determine preferred interpretation?
  - All relevant information and constraints use immediately (not just syntax)

- Linking Hypothesis:
  - Comprehension is easy when constraints support a common interpretation, and difficult when they conflict/compete

The Competitive-Integration Model

- **Claim**: Diverse constraints (linguistic and conceptual) are brought to bear simultaneously in ambiguity resolution.

- **The Model**: Assumes the all analyses are constructed
  - Constraints provide “probabilistic” support for each analyses
    - Constraint are weighted and normalized
    - Lexical & structural bias, parafoveal cues, thematic fit, discourse context ...

- **Goal**: Simulate reading times
  - RTs are claimed to correlate with the number of cycles required to settle on one of the alternatives

(McRae et al., 1998; Tanenhaus et al., 2000)
The Computational Model

The crook arrested by the detective was guilty of taking bribes

1. Combines constraints as they become available in the input

2. Input determines the probabilistic activation of each constraint

3. Constraints are weighted according to their strength

4. Alternative interpretations compete to a criterion

5. Cycles of competition mapped to reading times

Steps in the Experiment: (McRae et al 1998)

Constraints contribute to the activation of competing analyses, over time

1. Identifying the relevant constraints

2. Computational model for the interaction of constraints

3. Estimate bias of each constraint from corpora & rating studies

4. Weight of each constraint: fit with off-line completions

5. Make predictions for reading times

6. Compare actual reading times with those of:
   - Constraint-based model
   - Garden-path model
Constraint Parameters

“The crook/cop arrested by the detective was guilty of taking bribes”

Verb tense/voice constraint: is the verb preferentially a past tense (i.e. main clause) or past participle (reduced relative)

Relative log frequency is estimated from corpora: RR=.67 MC=.33

Main clause bias: general bias for structure of “NP verb+ed …”

Corpus: P(RR|NP + verb-ed) = .08, P(MC|NP + verb-ed) = .92

by-Constraint: extent to which ‘by’ supports the passive construction

Estimated for the 40 verbs from WSJ/Brown: RR= .8 MC= .2

Thematic fit: the plausibility of crook/cop as an agent or patient.

Estimated using a rating study

by-Agent thematic fit: good Agent is further support for the RR vs. MC.

Same method as (4).

Thematic Fit Parameters

“How common is it for a

crook _____
cop _____
detective _____
police _____
suspect _____

To arrest someone?

To be arrested by someone?
The Computational Model

The crook arrested by the detective was guilty of taking bribes

1. Combines constraints as they become available in the input
2. **Input determines the probabilistic activation of each constraint**
3. Constraints are weighted according to their strength
4. Alternative interpretations compete to a criterion
5. Cycles of competition mapped to reading times

The recurrence mechanism

- $S_{c,a}$ is the raw activation of the node for the $c^{th}$ constraint, supporting the $a^{th}$ interpretation,
- $w_c$ is the weight of the $c^{th}$ constraint
- $I_a$ is the activation of the $a^{th}$ interpretation
- 3-step normalized recurrence mechanism:
  1. Normalize: $S_{c,a}(norm) = \frac{S_{c,a}}{\sum_a S_{c,a}}$
  2. Integrate: $I_a = \sum_c [w_c \cdot S_{c,a}(norm)]$
  3. Feedback: $S_{c,a} = S_{c,a}(norm) + I_a \cdot w_c \cdot S_{c,a}(norm)$
A Gated Completion Study

- Establish that thematic fit does in fact influence “off-line” completion
- Use to adjust the model weights
- Manipulated the fit of NP1:
  - Good agents (and atypical patients)
  - Good patients (and atypical agents)
- Hypotheses: Effect of fit at verb
  - Additional effect at ‘by’
  - Ceiling effect after agent NP

Gated sentence completion study:

- The {cop,crook} arrested ...
- The crook arrested by ...
- The crook arrested by the ...
- The crook arrested by the detective...

Fitting Constraint Weights

- Adjust the weights to fit “off-line” data:
  - Brute force search of weights (~1M)
  - 20-40 cycles (step 2)
  - Node activation predicts proportion of completions for each interpretation
  - Weights determined by averaging the 10 best models from each of 20-40 cycles (110 models in total)

Counted “the crook arrested himself” as RR (!?)
The Complete Model

The crook arrested by the detective was guilty of taking bribes

1. Combines constraints as they become available in the input
2. Input determines the probabilistic activation of each constraint
3. Constraints are weighted according to their strength
4. Alternative interpretations compete to a criterion
5. Cycles of competition mapped to reading times

On-line study

- Two-word, self-paced presentation: (similar to completion studies)

The cop / that was / arrested by / the detective / was guilty / of taking bribes [GA,UR]
The cop / arrested by / the detective / was guilty / of taking bribes [GA,R]
The crook / that was / arrested by / the detective / was guilty / of taking bribes [GP,UR]
The crook / arrested by / the detective / was guilty / of taking bribes [GP,R]

Constraint Based (CB) Model
MC bias: .5094 x .75
Thematic Fit: .3684 x .75
Verb tense: .1222 x .75
by-bias: .25
Model Predictions

- Two “Versions” of the models:
  - **Constraint-Based**: constraints apply immediately for each region
  - **Garden-Path**: MC-bias & Main-Verb bias only, other constraints delayed one “region”

- Prediction Per-Region Reading times for each model:
  - Each region is processed until it reaches a (dynamic) criterion:
    \[
    \text{dynamic criterion} = 1 - \Delta \text{crit} \times \text{cycle}
    \]
  - As more cycles are computed, threshold is relaxed
  - \(\Delta \text{crit} = .01\) means a maximum of 50 cycles

“No model-independent signature data pattern can provide definitive evidence concerning when information is used”

CB vs. GP Model Predictions

- Constraint Based (CB) Model
  - MC bias: .5094 x .75
  - Thematic Fit: .3684 x .75
  - Verb tense: .1222 x .75
  - by-bias: .25

- Garden Path (GP) Model:
  - MC bias: 1

---

**FIG. 3.** Self-paced predictions derived from the constraint-based competition model. In this and all following model figures, the number beside each model datum is the mean activation of the reduced relative node after competition in that region for either (good agents) or (good patients).

**FIG. 4.** Self-paced predictions as derived from the garden-path model when constraints other than the main clause and main verb frames were delayed by a region.
The GP-model, has a 1-2 word delay in use of information, what if this delay is reduced? 4 cycles (10-25ms)

Better fit, but high reduction effect still predicted at main verb (good patient).

Search for the (new) best weights:

- MC bias: .2966 (.5094)
- Thematic fit: .4611 (.3684)
- V.tense: .0254
- by-bias: .2199

No-longer models completions
Modeling discourse constraints

1 Referent: An actress and the producer’s niece were auditioning for a play. The director selected the actress but not the niece. **The actress selected by the director believed that her performance was perfect.**

2 Referents: Two actresses were auditioning for a play. The director selected one actress but not the other. **The actress selected by the director believed that her performance was perfect.**

Thematic fit of actress doesn’t help much, but the discourse context does:

- Two referent context increases the likelihood of a relative clause

Revised model
Interpretation of by-phrases

Agent context: The artist decided to go to the gallery. Once he got there he wanted to know who had hung his prize painting. He was pleased to discover that his painting had been hung by the director earlier in the week.

Location context: The artist decided to go to the gallery. Once he got there he wanted to know where his prize painting had been hung. He was pleased to discover that his painting had been hung by the entrance earlier in the week.

• Relevant constraints: preference due to the verb, general interpretation of by-phrases in passives, and the discourse context

Results
Summary of McRae et al 1998

- **Interactive model** of constraint-based ambiguity resolution

- **Quantitative linking hypothesis** relating constraint-integration and reading times

- An **empirical evaluation** which includes:
  - setting constraint biases (using corpus estimates, and human judgements)
  - setting model weights (to fit completion data)
  - evaluating the predictions (against human reading times)
  - comparing constraint-based and modular predictions

Issues and Criticisms

- What “constraints” to include/exclude:
  - Ok if materials don’t vary w.r.t excluded constraint, or if excluded constraint correlates with (thus captured by) an included constraint

- Separate models for each phenomena: lack of generalization?

- Models constraint integration independent of parsing?
  - What is *really* being modeled? Can the approach scale?

- Predicts long reading times when constraints compete
  - Problem: People are sometimes *faster* at processing ambiguous regions!
Readings so far …


