Computational Psycholinguistics

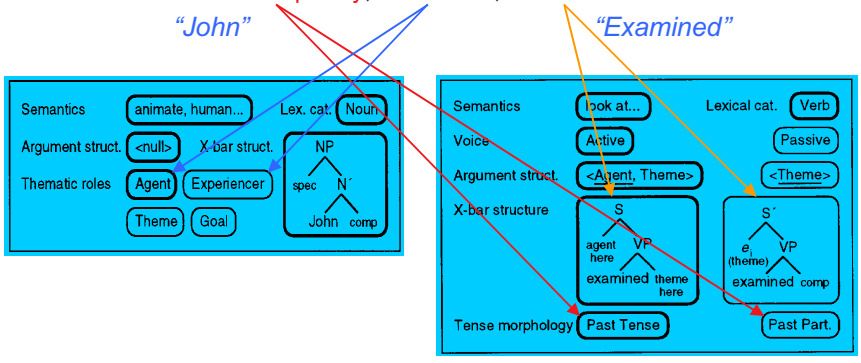
Lecture 10: Kinds of Models

Matthew W Crocker

Computerlinguistik
Universität des Saarlandes

The Interactive Activation Model

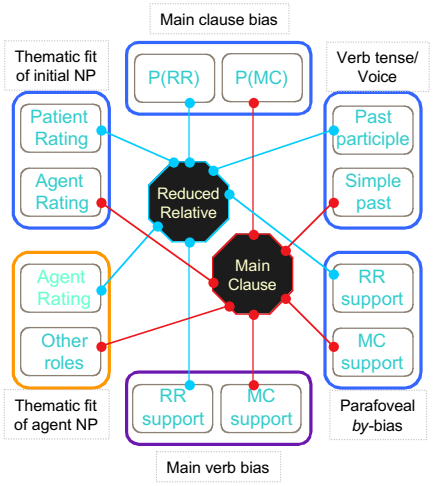
- Rich syntactic/thematic features
- Frequency determines 'activations'
- Consider: *"John examined the evidence"*
  - "examined" is ambiguous, as either a simple past or past participle
  - ➡ Constraints: tense frequency, thematic fit, structural bias ...



The Competition-Integration 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



Summary of the model

- Contraint biases set by off-line data:
  - Corpus frequencies
  - Norming studies
- Weights for constraints set by modelling a gating experiment
  - I.e. also off-line
- On-line (self-paced) data modelled by the number of cycles to reach criterion, in each region
- Better modelling of data when
  - All constraints are used immediately
- Than when
  - Only MC bias is used: “GP Model”
  - MC immediate, other constraints delayed one region: “GP Delay model”
- Is this a fair “replication” of other theories?

Issues and Criticisms

- Decision about what constraints to include/exclude, McRae et al:
  - Less important if materials don't vary w.r.t excluded constraint, or,
  - Of bias of excluded constraint correlates well with included constraint:
    - E.g. tense bias (included) correlates well with transitivity (excluded)
- Not a model of language *processing*:
  - Is it legitimate to characterise information flow separate from the structure building mechanism.
  - What is *really* being modelled? Can the approach be scaled up?
- Garden-path: A straw man
  - Is the implementation of the GP model fair, for purposes of comparison
  - What other constraints might be considered purely syntactic.
- Predicts long reading times when constraints are in close competition
  - In fact, people are often *faster* at processing ambiguous regions!
- Not truly probabilistic: activations only *begin* as probabilities
  - Also, many probabilities are derived from ratings (not frequencies)

Hybrid Models of Sentence Processing

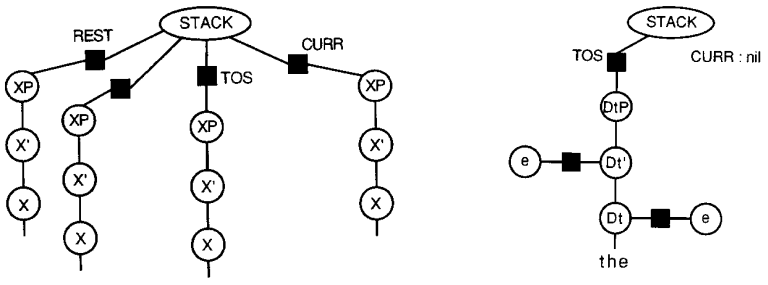
- Is it possible to construct competitive, activation-based models of the human parsing mechanism?
- Hybrid-models: exploit distributed computation, and competitive-activation with models of symbolic parsing:
  - TSVB Parser (Henderson): Embeds an implementation of Marcus' D-Theory in a connectionist architecture
  - CAPERS (Stevenson): Combines symbolic representation of structures, with attachments implemented as "connections"
  - Unification-Space (Kempen & Vosse): similar to Stevenson, but uses a different, stochastic mechanism

The Competitive Attachment Model

- CAPERS:
  - A hybrid symbolic-connectionist model of human sentence parsing
  - A competition-based model of parsing and reanalysis
- Models parsing, disambiguation, and reanalysis via competitive activation among structural alternatives
  - Direct, symbolic encoding of linguistic representations
  - Distributed decision-making
  - Competition-based spreading activation
    - Does not rely on inhibition
    - Competition is indirect, nodes vie for out from common neighbour

Building blocks of CAM

- Words instantiate X' projection templates
- Lexical item determine valency of projections:
  - specifiers, complements, and modifiers

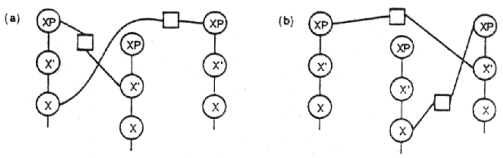


Implementation of the Model

- Nodes in the tree correspond to p-nodes, and are only projected on the basis of lexical input.
- Attachments between sisters are formed by a-nodes:
  - Mediate feature agreement between p-nodes
  - Each p-node uses constraint-based spreading activation (CBSA) to allocate activation to it's a-nodes:
    - Proportional to the current activation of the a-nodes
  - The degree of satisfaction of grammatical constraints determines the a-nodes state-value, which in turn contributes to the activation
  - A-nodes AND their inputs, to ensure that they "agree"
  - Null (phi-nodes) are inserted for attachments which are yet to be made
- All XPs must activate exactly one a-node,
  - since they must be attached to exactly one mother
- For other nodes, it depends on the complements or specifiers licensed by the lexical items

Restrictions on the Model

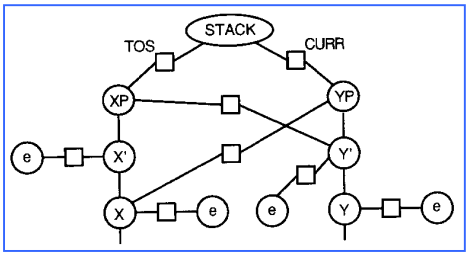
- If all possible attachments are established between the current phrase and the existing network, CBSA cannot ensure a consistent parse:
  - CBSA can only ensure 'locally' grammatical activation of a-nodes
  - Cannot rule out simultaneous activation of all inconsistent attachments



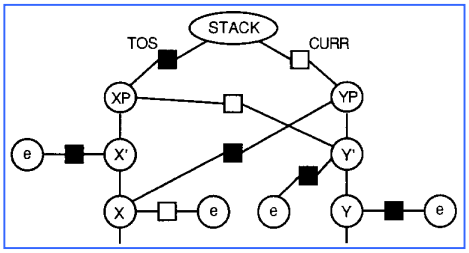
- Solution: Current phrase can only attach to the right edge of the tree
  - If no attachment is possible, attach to stack
    - Dispreferred, since no grammatical constraints are satisfied
  - Losing a-nodes are always deleted from the network
    - Avoids having them reactivate for form inconsistent parse trees
    - System always has only a single parse state (serial)

The Attachment Space

- Two projections may attach to each in 3 ways:

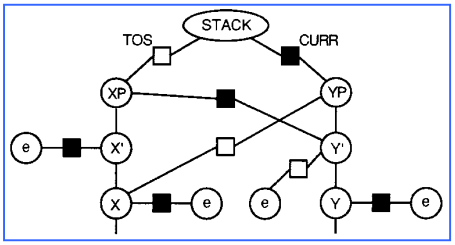


1. YP as complement of XP:

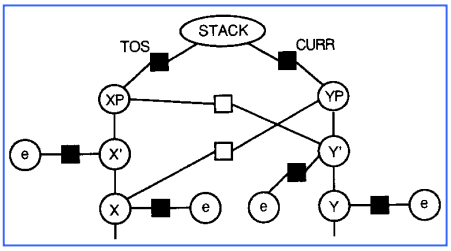


Attachment Space (cont'd)

2. XP as specifier of YP:

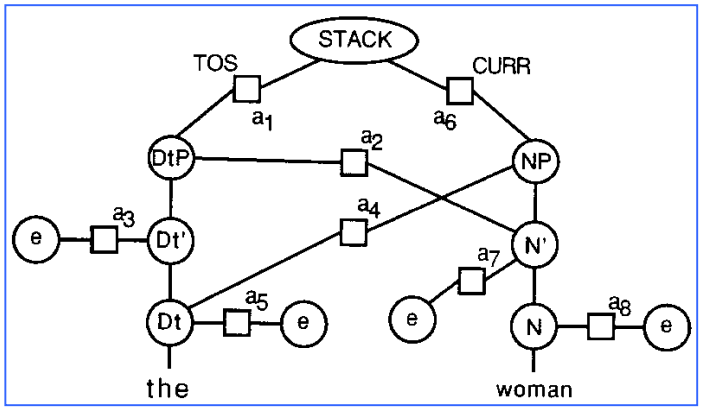


3. Or no attachment:



For example ...

- *"The woman …"*



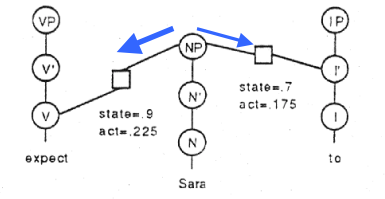
CBSA

- The CBSA Function
  - o_{ji}: output from n_i to n_j
  - a_i: activation of n_i
  - k: ranges over nodes connected to n_i

$$o_{ji} = \frac{a_j}{\sum_k a_k} \bullet a_i$$

- Consider: *Mary expected Sarah to leave*

- A-nodes "state" reflects degree to which grammar constraints are satisfied
- The output activation of p-nodes:
  - Shared to it's a-nodes, proportional to their current activation



Examples in detail

- Mary expected Sara to leave
- Sara:
  - Default stack attachment is not competitive, since a1 is highly activated
- to:
  - IP is projected
  - a2, a3 simultaneously present a revised structure possibility
  - a1 competes with a2, a3 for activation from the V & NP nodes
  - IP and I' can put all activation into a2, a3 respectively

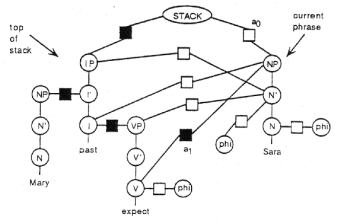
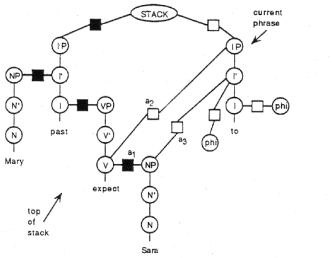


Figure 7: The network after attaching the NP *Sara*.



When Kiva eats food gets thrown

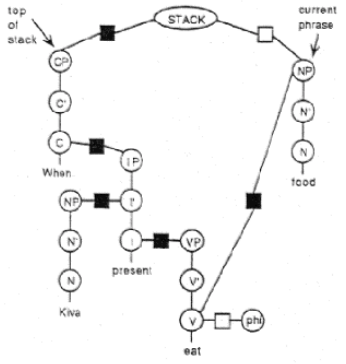
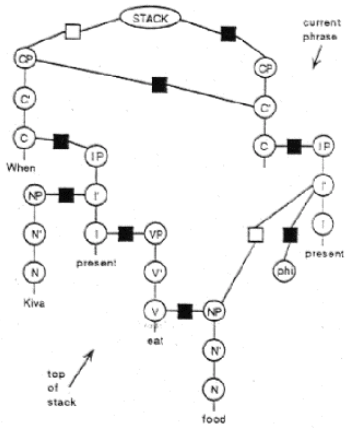
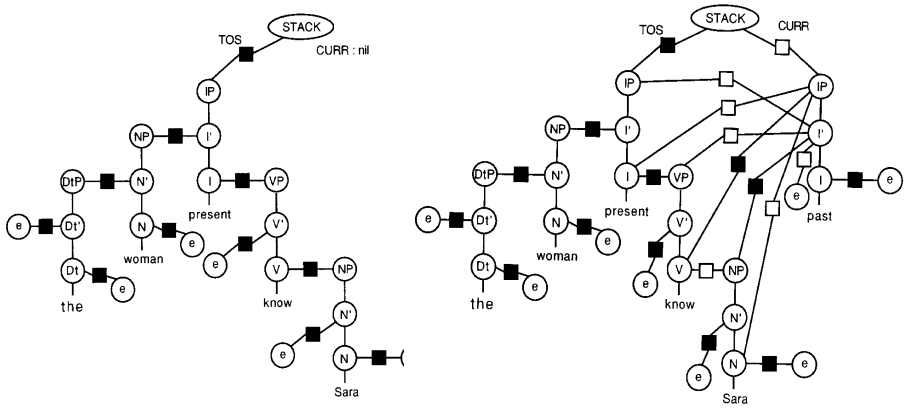


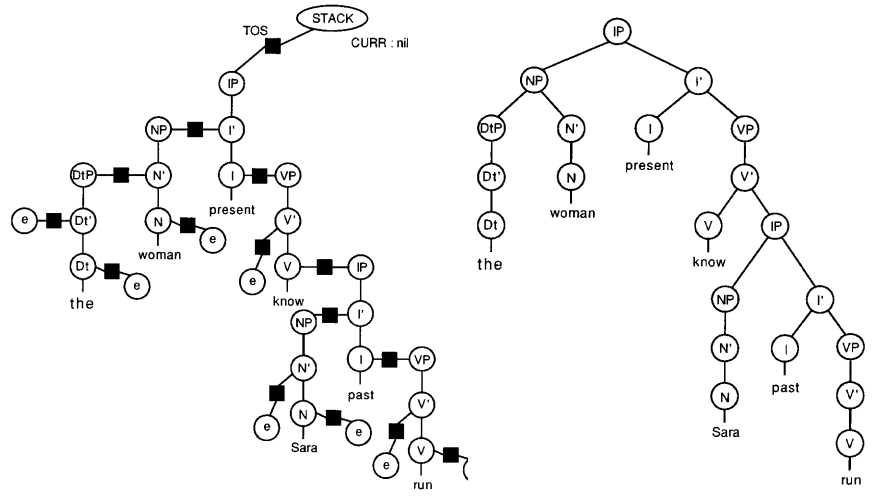
Figure 10: The NP *food* has a single valid attachment to the parse tree.



Stevenson's Model: Example 1

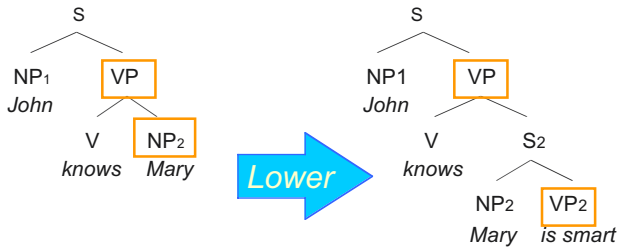


Example 1: continued



Monotonic Parsing Operation

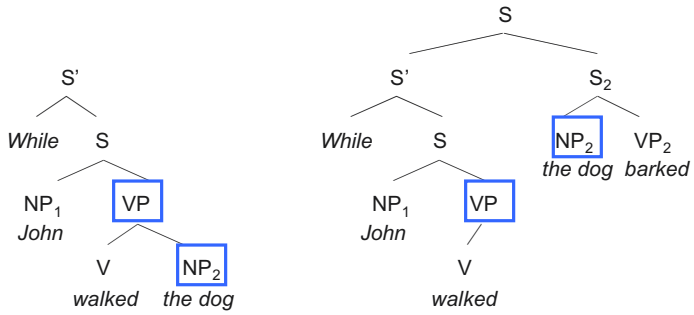
- Tree-lowering
  - John knows Mary … John knows Mary is smart



  - {dom(S,NP1), dom(S,VP), dom(S,V), dom(S,NP2), prec(NP1,VP), dom(VP,V), dom(VP,NP2), prec(V,NP2), dom(VP,S2, dom(S2,NP2), prec(NP2,VP2) …}

Non-Monotonic Parsing

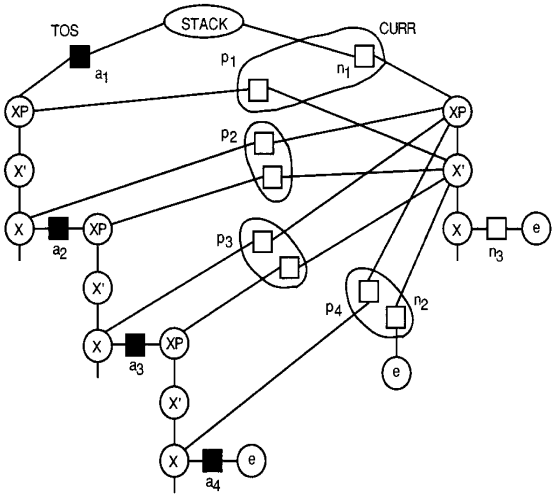
- Predicting difficult reanalysis
  - While John walked the dog … barked.



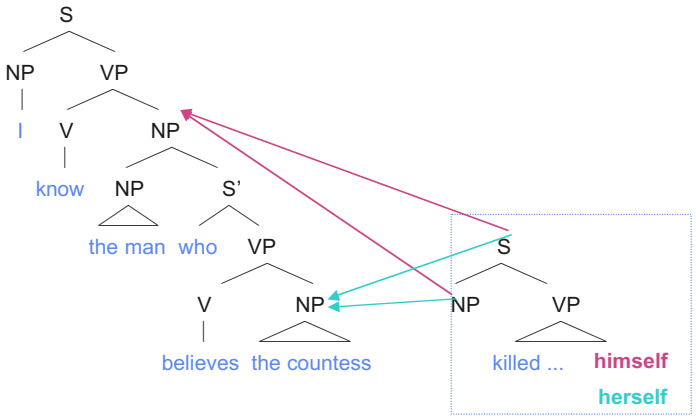
  - … *dom*(VP,NP_2) … ***but*** … *dom*(VP,NP_2) & *prec*(VP,NP_2) …

Three-site Competition

- Either: $a_1..a_4$ stay active and so $n_1..n_3$ activate
- Or: one of a_i become inactive, and the p_i pair must activate



The Trees



Summary of the Models Behaviour

- Attachment activations are driven by maximising the satisfaction of grammatical constraints (*a la* Pritchett)
  - After competition, losing attachments are removed
- Reanalysis is only possible when new attachments are strong enough to break existing ones, otherwise, garden-paths are predicted
  - NP/S complement reanalysis is possible
    - The woman knows Sarah runs
  - NP/Z complement is a garden path
    - When Kiva eats food gets thrown
    - Why? Eats has no other node to direct its activation to, so won't let go!
  - Also:
    - Jamie gave the child the dog bit a bandage
    - I convinced her children are noisy
- Behaviour is similar to Sturt & Crocker, but explanation is different
- Currently no mechanism for the interaction of other constraints

New Evidence for Modularity

- Interactive models rely on multiple analyses:
  - These is evidence about the preferred structure
  - No evidence for the existence of alternatives
- Syntactic preferences can't be reversed (TTK?)
  - I.e. in RR/MC ambiguity, the MC is never a garden path
- Neuroscientific data: against homogeneity
  - P600 (syntactic) vs. N400 (semantic) ERPs
- Localisation: Aphasia, PET scan