Human Sentence Parsing

• We are looking for parsing methods that
  • process the input sentence incrementally
  • construct a connected structure (interpretation)
  • explain experimental findings on human subjects, especially when facing ambiguity
Ambiguity in Parsing

• Ambiguity: what if at some point in parsing, more than one choice can be made?

• Parsing paradigms w.r.t ambiguity

  • **Serial parsing**: maintain a single parse at a time
    • Backtracking: go back to a choice point and select a different route
    • Determinism: use additional information to make the best decision

  • **Parallel parsing**: pursue many possible parses at the same time
Backtracking

**Method:**

- When more than one structure can be built, choose one and mark it as a choice point.
- If the analysis cannot be completed, undo everything up to a choice point, and select a different route.

**Selecting a choice point: the last one?**

- This involves the least effort, since a smaller portion of the sentence is re-parsed.
- This is intuitive, since earlier mistakes would have been discovered sooner.
Consistent with Human Parsing?

- Re-considering a more recent choice point means less processing time

The man saw the book was open.  

The woman sent the letter was pleased.
The man saw the book was open. After John left the shop closed.
Strategies for Disambiguation

- Choosing a random rule and backtracking to the last choice point does not explain some of the experimental findings

- Do humans use specific strategies when encountering ambiguity?
  - **Structural** strategies
  - **Grammatical** strategies
  - **Experiment-based** strategies
Structural Strategies

- Frazier (1979): **The Garden Path Theory**
  - Syntactic processor is guided by structural principles
- Minimal Attachment (MA)
  - Attach incoming material into the structure using the fewest nodes possible
- Late Closure (LC)
  - When possible, attach incoming material into the clause or phrase currently being parsed
John saw the man with the telescope.
NP/S Complement Ambiguity:

The student knew the solution to the problem.
The student knew the solution was incorrect.
Late Closure (LC)

- When MA does not solve the problem...

   The reporter said the plain crashed last night.
The Garden Path Theory

- Incremental, left-to-right, serial parsing
- When facing an ambiguity, use MA and LC to determine the parser’s decision
  - MA has a higher priority
- If the analysis later turns to be incorrect, a garden-path occurs => backtrack and re-analyse
  - Both conscious (hard to recover from) and unconscious (increasing complexity) garden paths are considered
Accounting for Human Preference

NP/VP Attachment Ambiguity:
“The cop [saw [the burglar] [with the binoculars]]”
“The cop saw [the burglar [with the gun]]”

NP/S Complement Attachment Ambiguity:
“The athlete [realised [his goals]] last week”
“The athlete realised [[his goals] were unattainable]”

Clause-boundary Ambiguity:
“Since Jay always [jogs [a mile]] [the race doesn’t seem very long]”
“Since Jay always jogs [[a mile] doesn’t seem very long]”

Reduced Relative-Main Clause Ambiguity:
“[The woman [delivered the junkmail on Thursdays]]”
“[[The woman [delivered the junkmail]] threw it away]”

Relative/Complement Clause Ambiguity:
“The doctor [told [the woman] [that he was in love with her]]”
“The doctor [told [the woman [that he was in love with]] [to leave]]”
Summary of Frazer (1979)

• Strategies are defined in terms of the form of syntactic structure (e.g., number of nodes)

• Psychological assumptions:
  • **Modularity**: only syntactic information used for initial structure building
  • **Resources**: emphasizes importance of memory limitations
  • Processing strategies are universal and innate
Grammar-based Strategies

- Strategies are based on the content of the syntactic structures
  - Various syntactic positions are distinguished with respect to their function and content
- Strategies are derived from the purpose of the task, not the computational efficiency
  - Less sensitive to minor structural details
Pritchett (1988)

- Incrementally satisfying various syntactic constraints
- Based on Chomsky’s Government-binding theory
  - Each verb has a number of thematic roles which must be satisfied (e.g., Agent, Theme, Destination...)

He is mending a sock on the table.
Pritchett (1988)

- Assumption: lexical entries of verbs are fully specified for thematic roles

- **Theta-Attachment**:
  - Maximally satisfy the theta-criterion at every point during processing

- **Theta Reanalysis Constraint**:
  - Reanalysis of a constituent out of its theta-domain results in conscious garden-path
While Mary was mending the sock it fell off her lap.

While Mary was mending the sock fell off her lap.
While Mary was [vp mending [np the sock]] [s it fell off her lap].

But mend needs a Theme

While Mary was [vp mending] [s [np the sock] fell off her lap].
Reanalysis to a position **within** the original theta domain is easy:

```
S ei NP                 VP ei NP
6       ru g k knew
NP                VP
6        6
the solution was incorrect
```

```
The student V                         S
S ei NP                 VP ei NP
6       3 g ro knew
NP                VP
6        6
the solution
```
• Reanalysis to a position outside the original theta domain is difficult:
Another Example

Without her contributions the orphanage closed.

*Without*: a preposition with a single thematic role

*her*: a determiner of an unseen NP head, or a full NP (pronoun)

*contributions*: head of a new NP with no role, or combine with her for a full NP

Theta Attachment

Without her contributions failed to come in.

*contributions*: becomes subject of *failed* violating Theta Reanalysis Constraint
Accounting for Human Preference

NP/VP Attachment Ambiguity:
“The cop [saw [the burglar] [with the binoculars]]”
“The cop saw [the burglar [with the gun]]”

NP/S Complement Attachment Ambiguity:
“The athlete [realised [his goals]] last week”
“The athlete realised [[his goals] were unattainable]”

Clause-boundary Ambiguity:
“Since Jay always [jogs [a mile]] [the race doesn’t seem very long]”
“Since Jay always jogs [[a mile] doesn’t seem very long]”

Reduced Relative-Main Clause Ambiguity:
“[The woman [delivered the junkmail on Thursdays]]”
“[[The woman [delivered the junkmail]] threw it away]”

Relative/Complement Clause Ambiguity:
“The doctor [told [the woman] [that he was in love with her]]”
“The doctor [told [the woman [that he was in love with]] [to leave]]”
Grammar-based Strategies

- Theta Attachment: reliance on verb thematic information means that it is head driven
  - Problematic for verb-final languages
- Crocker (1992): Argument Attachment
  - Attach constituent into potentially role-receiving positions
Theories of Sentence Processing

• Theories of parsing typically determine …
  • what **architecture** is assumed: modular? symbolic? …
  • what **mechanism** is used to construct interpretations?
  • which **information** sources are used by mechanism?

• **Linking Hypothesis**: Relate theory/model to observed measures
  • Preferred structures should have faster reading times in the disambiguating region than dispreferred
Frazier’s Garden-Path Theory

- **Architecture**: modular syntactic processor, with restricted lexical (category) and semantic knowledge

- **Mechanism**: incremental, serial parsing, with reanalysis

- **Information**: general syntactic principles based on the current phrase structure

- **Linking Hypothesis**:
  - Parse complexity and reanalysis cause increased RTs
Pritchett’s Theory

• **Architecture**: modular lexico-syntactic processor with syntactic and thematic role features

• **Mechanism**: incremental, serial parsing, with reanalysis

• **Information**: grammar principles and thematic role information

• **Linking Hypothesis**:
  
  • TRC violation causes garden-path, reanalysis without TRC is relatively easy