The Role of the Comma in the Resolution of Closure Ambiguities
Vered Argaman and Aurora Mendelsohn

Introduction

Subordinate / Main Clause Ambiguity
After Janet left the party she ran into some old friends. Late Closure (LC)
After Janet left the party began. Early Closure (EC)

Several areas of research have relied heavily on empirical evidence from this ambiguity.
• Late Closure Principle in Garden Path (Frazier & Rayner, 1982)
• Thematic role retention (Christianson, Hollingworth & Ferriera, 1999)
• Factors that affect ease of reanalysis (Ferriera & Henderson, 1991; Sturt, et. al, 1999)
In all cases, all ambiguous versions are presented without commas.

Assumptions about the Validity of the Comma-less Versions
It has been assumed that parsing strategies can function independently of commas (Frazier & Rayner, 1982) because:
• Not all ambiguities can be helped by a comma (e.g. MV/RR, VP-Attachment) and so commas are not reliable cues for disambiguation
• Though required by prescriptive rules, commas are used inconsistently in real production

A Theoretical Alternative to Late Closure: Constraint-based Models
• LC is preferred because the parser is sensitive to the relative frequencies of alternatives
This model can be extended to include sensitivity to commas as an additional source of information that constrains choices.
• If commas are usually present, the absence of comma in the EC interpretation could strengthen the bias towards the LC interpretation.
• Commas are an effective disambiguating cue in this ambiguity; readers slow down while reading the comma but have no trouble at disambiguation (Hill & Murray, 1997, 1998).
• Sensitivity to both overall and conditional probabilities of comma use

Are commas in these sentences truly optional? - Experiment 1
Do commas constrain parser choices? - Experiment 2

<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants completed 12 sentence-initial fragments consisting of a subordinating conjunction followed by a proper name</td>
</tr>
<tr>
<td><strong>While Lorna</strong></td>
</tr>
</tbody>
</table>

Subordinating Conjunctions:
*after, although, as, because, even though, if, in case, once, though, since, when, while*

• 12 fillers: adverbs followed by proper name, e.g., “Yesterday Denise…”
• 140 participants

Coding
- closure (early vs. late)
- three possible disambiguating cues:
  - comma usage
  - embedded verb transitivity
  - case marking of the NP immediately following the verb

Experiment 1 Results

<table>
<thead>
<tr>
<th>Token Frequency by Closure and Comma Use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comma</strong></td>
</tr>
<tr>
<td>Early</td>
</tr>
<tr>
<td>194 (30%)</td>
</tr>
<tr>
<td><strong>No Comma</strong></td>
</tr>
<tr>
<td>Early</td>
</tr>
<tr>
<td>53 (8%)</td>
</tr>
<tr>
<td>247 (38%)</td>
</tr>
</tbody>
</table>

• Commas used in 78% of the sentences in both EC and LC
• Of 141 sentences without a comma, 65% had at least one other cue, so overall only 8% (50) of all the sentences were ambiguous

The ambiguous sentences (with none of the three possible disambiguating cues) were mainly LC (84%) and rarely produced a noticeable garden path effect because many were disambiguated by plausibility.

Conclusions
• Commas are used consistently in these sentences.
• LC is the preferred structure in natural production.
• Truly ambiguous versions of this structure are rare.
This raises serious questions for research based on the assumption that the comma is optional and that the parser is unperturbed by its absence.
Experiment 2: Comma Tracking

- Different subordinating conjunctions from Experiment 1 had different frequencies of taking a comma, from 67% to 91%.

Prediction: If readers are sensitive to these conditional probabilities, then the more often a conjunction appears with a comma, the greater the difficulty will be in its absence.

<table>
<thead>
<tr>
<th>Subordinating Conjunction Frequencies</th>
<th>High Frequency % Comma</th>
<th>Low Frequency % Comma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Even though</td>
<td>90.9</td>
<td>67.3</td>
</tr>
<tr>
<td>As</td>
<td>83.3</td>
<td>71.8</td>
</tr>
<tr>
<td>In case</td>
<td>84.1</td>
<td>70.8</td>
</tr>
<tr>
<td>Mean</td>
<td>86.1</td>
<td>70.0</td>
</tr>
</tbody>
</table>

Experiment: Manipulate Ambiguity and Comma Frequency

High Comma Frequency - Ambiguous

Even though the janitor vacuumed, the carpet was covered with dust and crumbs.

Low Comma Frequency - Ambiguous

When the janitor vacuumed, the carpet was covered with dust and crumbs.

High Comma Frequency - Unambiguous

Even though the janitor vacuumed, the carpet was covered with dust and crumbs.

Low Comma Frequency - Unambiguous

When the janitor vacuumed, the carpet was covered with dust and crumbs.

Method

- 24 critical items, all EC
- Fillers included LC items, with and without commas
- Self-paced, word-by-word, moving window; comprehension questions after each item
- Each conjunction was used 8 times and each combination of high and low frequency conjunctions was used 2-3 times
- 60 participants

Experiment 2 Results

- At disambiguating region (indicated by arrow):
  - Main effect of Ambiguity: standard garden path effect
  - Main effect of Comma Frequency
  - Interaction: higher comma frequency associated with greater ambiguity effect

Correlations with Ambiguity Effect:

- Ambiguity effect correlated with frequency with which each subordinating conjunction took a comma in Experiment 1 ($r = .91$)

- Other variables were calculated from the Wall Street Journal corpus:
  - Overall frequencies of each conjunction
  - Coded the first 100 tokens of each conjunction to obtain frequencies and conditional probabilities for 1) use as a subordinating conjunction and 2) place in the sentence
  - Ambiguity effect did not correlate with any of these variables

Conclusions

- Commas are tracked by the parser and used to guide parsing decisions.
- Both the Late Closure Principle and Constraint-based models can account for the ambiguity effect.
- Constraint-based models can easily account for the interaction between ambiguity and comma frequency.

Acknowledgement

This research was supported by the National Science Foundation (Grant # SBR-9729007).