A Puzzle Regarding Relative Clauses: When Frequency and Difficulty Disagree
Gabriel Doyle and Roger Levy – University of California, San Diego

Abstract
We consider an alternation seemingly without semantic or structural differences. The lack of such differences suggests that frequency should drive differential comprehension, but we show that observed differences in reading times point opposite the frequency effects.

Background
- Comprehension difficulty in relative clauses (RCs) affected by composition of the relative clause.
- Gordon et al (2001) showed similarity-based interference; similar extracted & embedded NP types impede comprehension.
- Traxler et al (2002) showed crucial role of plausibility; similar extracted & embedded NP plausibilities impede comprehension.
- Reali & Christiansen (2007) showed frequency of exposure of different RC types influences comprehension difficulty, independently of complexity or cognitive limitations.
- Corpus frequency and comprehension difficulty pattern together; surprisal theory (Hale 2001, Levy 2008).

How sensitive are comprehenders to word/structure cooccurrences that are purely arbitrary (not semantically based)?

Linking Hypotheses
- Dependency Locality Theory
- Filler-gap retrieval
- Similarity-based interference
- Surprise

Our Alternation
Restrictive relative clauses with an animate extracted NP have two possible relative pronouns:

1a) The man that I saw smiled.
1b) The man who I saw smiled.
- Structural equivalence
- Approx. semantic equivalence
- No apparent discourse differences
- Only differ in surprisal linking hypothesis

Without obvious difference in complexity, we expect frequency effects to determine comprehension difficulty.

Experimental Design
2x2x2 design:
- Relativizer (that/who)
- Extraction Type (Subject/Object)
- Embedded NP Type (Pronoun/Full NP)

Stimuli:
- [ORC-pro] The chef [that/who] you watched was famous for her butternut squash soup.
- [SRC-pro] The chef [that/who] you watched was famous for her butternut squash soup.
- [ORC-full] The chef [that/who] the waiter watched was famous for her butternut squash soup.
- [SRC-full] The chef [that/who] watched the waiter was famous for her butternut squash soup.

Experiment 1: Corpus Study
Combined the Brown, Switchboard, and Wall Street Journal corpora.
Performed tgrep searches to find restrictive RCs with animate extracted NPs.

<table>
<thead>
<tr>
<th>Relative Pronoun</th>
<th>ORC-pro</th>
<th>SRC-pro</th>
<th>ORC-full</th>
<th>SRC-full</th>
</tr>
</thead>
<tbody>
<tr>
<td>that</td>
<td>59</td>
<td>12</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>who</td>
<td>15</td>
<td>27</td>
<td>22</td>
<td>13</td>
</tr>
</tbody>
</table>

Fisher’s exact tests: significant differences between ORC-pro and SRC-pro RCs and between SRC-full and SRC-pro RCs.

Experiment 2: Reading-Time
We expect comprehenders to use this information in comprehension, and to find the same patterns in reading times.

Corpus frequency (Expt 1) predicts:
- interaction between ORC-pro and SRC-pro, ORC-pro favoring that.
- interaction between SRC-full and SRC-pro, SRC-full favoring that.
- corpus and RT bars should be reversed.

Moving window self-paced reading-time study with 56 participants.

Conclusions
- Comprehension difficulty in that/who alternation shows no effect of relative pronoun frequency.
- In fact, difficulty increases with increased corpus frequency.
- This despite apparent lack of other factors influencing comprehension difficulty.

Future Work
- Use the ANC corpus to obtain better ORC-full counts.
- Sentence completion task to examine production in a more controlled setting.

References

Contact
{gdoyle, rlevy}@ling.ucsd.edu