When can Resumptive Pronouns rescue islands?

This is an island that, although an RP seems to salvage [a/it], is violated.

Literature is split on the rescuing potential of Resumptive Pronouns (RPs): - RPs are attested in English corpora (e.g., Prince 1990) and produced in the lab (Ferreira & Swets 2005) ... - But are not rated higher than gaps in acceptability rating tasks (e.g., Ono & Nakao 2004, Alexopoulos & Keller 2007, Heestand et al. 2011, etc.) - English RPs are intrusive (Seis 1984) and ungrammatical (Kuzech 2004) ... - But are useful for processing complex structures, like islands and long-distance dependencies (e.g., Etschel-Schr 1992, Alexopoulos & Keller 2007)

Much psycholinguistics literature suggest RPs should be less degraded than gaps, but fail to support the suggestion experimentally (e.g., Alexopoulos & Keller 2007, Heestand et al. 2011, Han et al. 2012, Polinsky et al. 2013).

So if this difference has not been robustly observed in comprehension experiments, when can people use and make use of RPs?

We suggest the division in the literature is due to methodological differences, and find that a different paradigm demonstrates rescuing potential of RPs.

Background

Literature that supports rescuing potential evaluates differences between gap and RP with direct comparison (McDaniel & Cowart 1999) or production (corpus: Prince 1990; experiment: Ferreira & Swets 2005).

Literature that does not support rescuing potential evaluates differences between gap and RP on a scale that includes grammatical sentences: - Likert scales (Ono & Nakao 2004, Heestand et al. 2011, Polinsky et al. 2013)

- Magnitude estimation (Alexopoulos & Keller 2007, Han et al. 2012)

☆ Cf. comprehensibility scale, not acceptability (Beltrama & Xiang CUNY 2013)

Questions:
1) Could studies that fail to show amelioration with RPs be using the wrong type of baseline?
- If comparison set includes fully grammatical option, the RP and gap options may not be distinguishable
2) What kind of baseline is appropriate?
- Studies that show amelioration may explicitly limit evaluations to RP versus gap situations
3) Is the production - comprehension distinction relevant, or an artifact of the methodologies?
- Could production of RPs in islands have an implicit comparison set that only includes gap and RP options (not options without island violations)?

Methodology

Current study:
- Forced choice paradigm is similar to both comprehension and production:
  - Sentence completion is like production with very limited set of options
  - Full sentence choice is like comprehension (acceptability rating) with binary rating options

Participants:
- Amazon Mechanical Turk (MTurk)
- 80 unique participants (40 for each experiment)
- Self-reported native English speakers from USA

Experiment 1: Forced-choice sentence completion
a) Island Condition
Which man did Jane say that the parent who ___ forgave the babysitter’s mistake?
scolded
scolded him
b) Matrix Condition
Which man did Jane say that the parent who scolded the caregiver ___?
for-gave him

Experiment 2: Full-sentence forced-choice
- Sentences are the same as in (1), presented in pairs, without blanks:
  - Which man did Jane say that the parent who scolded forgave the babysitter’s mistake?
  - Which man did Jane say that the parent who scolded him forgave the babysitter’s mistake?

Predictions for the Island Condition
1a) If Gap is an appropriate baseline for direct comparison to RP
- RP >> Gap
McDaniel & Cowart 1999
1b) If Gap is not an appropriate baseline for direct comparison to RP
- RP >> Gap, no distinction is detectable
e.g., Heestand et al. 2011

2a) If RPs are only possible in production (for speaker, not listener)
- Experiment 1: RP >> Gap
Ferreira & Swets 2005
- Experiment 2: RP = Gap
e.g., Polinsky et al. 2013

2b) If RPs are useful in production and comprehension
- Experiment 1: RP >> Gap
Ferreira & Swets 2005
- Experiment 2: RP >> Gap
Beltrama & Xiang 2013

Results

Experiments 1 and 2:

<table>
<thead>
<tr>
<th>Island Condition</th>
<th>Matrix Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gap = 17.9%</td>
<td>Gap = 62.1%</td>
</tr>
<tr>
<td>RP = 82.1%</td>
<td>RP = 37.9%</td>
</tr>
<tr>
<td>Island</td>
<td>Matrix</td>
</tr>
<tr>
<td>RP = 82.1%</td>
<td>RP = 37.9%</td>
</tr>
</tbody>
</table>
| Use of RPs in Matrix and Island Positions
| Full-sentence rescued                                                                 |

![Use of RPs in Matrix and Island Positions](image)

<table>
<thead>
<tr>
<th>Island Condition</th>
<th>Matrix Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Island</td>
<td>Island</td>
</tr>
<tr>
<td>R = 82.1%</td>
<td>R = 80.0%</td>
</tr>
<tr>
<td>x² = 82.20, p &lt; 0.0001</td>
<td></td>
</tr>
<tr>
<td>95% CI: [74.3, 84.8]</td>
<td></td>
</tr>
<tr>
<td>Matrix</td>
<td>Matrix</td>
</tr>
<tr>
<td>R = 24.6%</td>
<td>R = 24.6%</td>
</tr>
<tr>
<td>x² = 61.00, p &lt; 0.0001</td>
<td></td>
</tr>
<tr>
<td>95% CI: [19.4, 30.6]</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

- Results are compatible with production and direct-comparison studies
- Forced-choice paradigm explicitly compares two structures known to be unacceptable in acceptability rating tasks
- Results are robust: RPs are preferred to gaps in islands
- Amelioration effect may only be detectable with comparison sets that do not include fully grammatical alternatives
- In conjunction with B&M13 and FBS05, this suggests speakers may use RPs to rescue islands in terms of comprehensibility, not acceptability

- Compatible with previous literature that doesn’t detect a difference between RP and gap

Future directions:
- Do other island types display similar effects?

References: