Relative Clause Association Ambiguity
(global structural ambiguity)

(1) 少女の兄
[isi-ga syokusinsiteiru] syôzyo-no ani
doctor-NOM palpating girl-GEN brother
[RC] N1 N2
(two association sites available after the RC)

‘the brother of the girl [(that) the doctor is palpating]’
NP1 NP2 [RC]
(attachment ambiguity: two attachment sites available prior to the RC)

Corpus Frequency of Relative Clause Association in Japanese
Toshiyuki YAMADA
Douglas ROLAND
Manabu ARAI
Yuki HIROSE
(The University of Tokyo)

Puzzle: Unforced Revision?
➢ If we believe in Revision as Last Resort (Fodor & Frazier, 1980),
the N1 association should be maintained.

Possibility: Structural Frequency Bias
➢ RC association ambiguity:

Roadmap
0. Overview
1. Introduction
2. Yamada et al. (2014)
3. The current study
4. Discussion
5. Conclusions

Structural Frequency Effects
➢ RC attachment ambiguity (Cuetos & Mitchell, 1988)
NP1 of-NP2 [RC]
- native English speakers: low attachment preference
- native Spanish speakers: high attachment preference
➢ Corpus frequency (Mitchell et al., 1992)
- English: 62% for low attachment ➔ low preferred
- Spanish: 60% for high attachment ➔ high preferred
➢ Frequency effects in PP attachment
(e.g., Katsika, 2011 for Greek, cf. Gibson et al., 1996 for English)
Previous Findings on RC Association

- Mixed results
  - N1 association preference (on-line evidence)
  - N2 association preference (off-line evidence)
  (e.g., Kamide & Mitchell, 1997; Miyamoto et al., 2004; Uetsuki, 2007; Nakano, 2008)
  a possibility: on-line revision
  N1 association \( \rightarrow \) N2 association

- Yamada et al. (2014): suggestive eye-tracking evidence
  (cf. Bai et al., 2014: no commitment to the N1 association)

On-line N2 Association Preference

(2a) N2-incompatible
[fai-ga tatta ima syokusinsiteiru] syôzyo-no ani-ga baien-de kaimonoseth. doctor-NOM just now palpating girl-GEN brother-NOM store-at shopping
‘The brother of the girl that the doctor is palpating just now is shopping at a store.’

(2b) N2-incompatible
[fai-ga tatta ima syokusinsiteiru] kanzya-no ani-ga baien-de kaimonoseth. patient-GEN brother-NOM store-at shopping
‘The brother of the patient that the doctor is palpating just now is shopping at a store.’

(2c) N2-compatible
[fai-ga tatta ima syokusinsiteiru] isu-de zittositeiru. chair-at sitting still
‘The brother of the girl that the doctor is palpating just now is sitting still on a chair.’

(2d) N2-compatible
[fai-ga tatta ima syokusinsiteiru] kanzya-no ani-ga isu-de zittositeiru. patient-GEN brother-NOM chair-at sitting still
‘The brother of the patient that the doctor is palpating just now is sitting still on a chair.’

Unforced Revision

- [RC]  N1-GEN  N2
  - initial commitment to the N1 association
  - revision to the N2 association

Although the N1 association is grammatical, Japanese comprehenders are willing to make an unforced revision to the N2 association at the point of N2.

- processing difficulty when the sentence-final main predicate is incompatible with the N2 association

Research Question 1

Is there a structural frequency bias towards the N2 association?

- Prediction
  a strong frequency bias towards the N2 association when N1 and N2 associations are both plausible, the situation most similar to the experimental items in Yamada et al. (2014)

Method

- corpus: Kyoto University Text Corpus
(http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?Kyoto%20University%20Text%20)
  • 38,400 samples (i.e., sentences) in total
- target samples: verb + N1 + no (GEN) + N2
- 4034 samples collected with restrictions, e.g.,
  1. verb dependent on either N1 or N2
  2. nouns could not be tame ‘for’, koto ‘the fact (that …)’, nado ‘etc.’, naka ‘inside’
- procedure: false positive matches were excluded
  • 861 samples remained for analysis

Data Analysis: Ambiguous RC Association

(5) Both N1 and N2 associations are plausible
… 抑収した山林の土 …
… osyuuusita sanrin-no tuti …
seized  forest-GEN soil
‘… the soil of the forest that was seized …’
3. The Current Study

Data Analysis: Forced RC Association

(6) a. N1 association forced
「オンリーワン」哲学を実践してきた三井氏の答え…
"Onrii wan" tetugaku-o zissensilekita Mitui-si-no kotae…
only one philosophy-ACC have.practiced M.-Mr.-GEN answer
'The answer of Mr. Mitsui that has practiced “Only one” philosophy…'

b. N2 association forced
低く抑えた男の声…
Hikuku osaeta otoko-no koe…
low made male-GEN voice
'The voice of the male that is made low…'

Data Analysis: N1 or N2 Association

(7) a. N1 association intended
この外交の意味と原則を理解しない政治家の言動…
Kono gaikou-no imi to gensoku-o rikaisinai seizika-no gendou…
this diplomacy-GEN meaning and principle-ACC understand.not politician-GEN behavior
'The behavior of the politician that does not understand the meaning and principle of this diplomacy…'

b. N2 association intended
時代の変化に合った事業の見直し…
Zidai-no henka ni atta zigyou-no minaosi…
times-GEN change-for suitable project-GEN revision
'the revision of the project that is suitable for the change accompanying the times…'

Frequency did not seem to explain RC association preferences in Japanese

<table>
<thead>
<tr>
<th>RC</th>
<th>N1 の</th>
<th>N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association</td>
<td>N1</td>
<td>N2</td>
</tr>
<tr>
<td>All examples</td>
<td>408</td>
<td>427</td>
</tr>
</tbody>
</table>

But, there was a N2 bias when only ambiguous examples were considered

<table>
<thead>
<tr>
<th>RC</th>
<th>N1 の</th>
<th>N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Association</td>
<td>N1</td>
<td>N2</td>
</tr>
<tr>
<td>Ambiguous examples only</td>
<td>237</td>
<td>336</td>
</tr>
</tbody>
</table>

Weak Bias towards the N2 Association

- Finding 1 no bias overall, but N2 bias when both associations were plausible
- Contrary to the prediction, the bias towards the N2 association was not sufficiently strong to override the semantic preference for N1 association in Yamada et al. (2014)
- The structural frequency bias is not the only source, but comprehenders may take ambiguity into account in generating association preferences

Research Question 2

Do we observe the N2 bias when the situation is most similar to the experimental items in Yamada et al. (2014)?

- Several interacting factors might affect comprehenders’ processing of [RC] N1 GEN N2
  - length of RC (short vs. long) (e.g., Hirose et al., 1998)
  - RC-type (subject- vs. object-extracted) (e.g., Uetsuki, 2006)
  - animacy of N1 and N2
  - properties of N1 and N2
    (i) discourse status and (ii) uniqueness
Possible Reasons for Modifying a Noun

1. Grounding a new referent (e.g., Fox & Thompson, 1990) → new referents are more likely to be modified

2. Identifying an old and non-unique referent → old and non-unique referents are more likely to be modified

(Two forces push in opposite directions, but they are reasonable ones in the discourse.)

Data Analysis: Discourse Status

- Criterion
  If the noun was either explicitly mentioned, or the existence of its referent was implied in the five preceding sentences, the N1/N2 was coded as discourse-old.

  - N1 coded as discourse-old
    - (8) "gakkou-o riyou ..." (preceding context)
      - school-ACC use
    - "syakaisihon-no ..." (the referent was implied)
      - social.capital-GEN

N2 more likely to be discourse-new

Suggestions that discourse-new referents are more likely to be modified

<table>
<thead>
<tr>
<th>Discourse Status</th>
<th>N1</th>
<th>N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>67</td>
<td>82</td>
</tr>
<tr>
<td>Old</td>
<td>33</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

But, discourse status did not affect RC association

Discourse status:
- ✓ N1-N2 modification
  - Only N1 new 4 5
  - Only N2 new 24 30
  - Both new 63 52
  - Both old 9 13

Research Question 2

Do we observe the N2 bias when the situation is most similar to the experimental items in Yamada et al. (2014)?

- Prediction
  - a frequency bias towards the N2 association when both referents of N1 and N2 are (i) discourse-new and (ii) non-unique
3. The Current Study

N2 more likely to be non-unique

Suggests that old and non-unique referents are more likely to be modified

<table>
<thead>
<tr>
<th>Uniqueness</th>
<th>N1</th>
<th>N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-unique</td>
<td>59</td>
<td>83</td>
</tr>
<tr>
<td>Unique</td>
<td>41</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

But, uniqueness did not affect RC association

Uniqueness: ✓ N1-N2 modification × RC association

<table>
<thead>
<tr>
<th>Uniqueness</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only N1 non-</td>
<td>Unique</td>
</tr>
<tr>
<td>unique</td>
<td>0</td>
</tr>
<tr>
<td>Only N2 non-</td>
<td>Unique</td>
</tr>
<tr>
<td>unique</td>
<td>35</td>
</tr>
<tr>
<td>Both non-</td>
<td>Unique</td>
</tr>
<tr>
<td>unique</td>
<td>59</td>
</tr>
<tr>
<td>Neither non-</td>
<td>Unique</td>
</tr>
<tr>
<td>unique</td>
<td>6</td>
</tr>
</tbody>
</table>

No Effects of N1/N2 Properties

- Finding 2: neither discourse status nor uniqueness affected RC association
- Contrary to the prediction, no bias towards the N2 association was found when both referents of N1 and N2 were (i) discourse-new and (ii) non-unique as in Yamada et al. (2014)
- While N1 may play a role in identifying N2 in the construction of [RC] N1-GEN N2, RC does not seem to play the same role

Summary

- Puzzle: unforced revision
  - [RC] N1-GEN N2

- N1 association → N2 association ...why?
- Answer: the frequency bias cannot account for the N2 association preference (Yamada et al., 2014)
  - weak bias towards the N2 association when both associations are semantically plausible
- Comprehenders may take ambiguity into account in generating RC association preferences

Future Issues

- We need a better understanding of the role of RC:
  - Why do speakers choose to modify either N1 or N2 with a RC?
- We need a better understanding of association biases in comprehension:
  - Do comprehenders really have a strong bias towards the N2 association?
  - Are the patterns in eye-movements in Yamada et al. (2014) due to other factors than the corpus frequency?

- It is still possible that distributional factors such as the following may lead to the N2 association bias:
  1. prosodic (phonetic-phonological) factors such as length of RC, length of N1 and N2 (e.g., Hirose et al., 1998; Nakano & Kahraman, 2013)
  2. morpho-syntactic factors such as RC types (i.e., subject/object-extracted), the matrix positions of N1 (i.e., subject/object), inherent adjunctness of N1 to N2 (e.g., Uetsuki, 2006)
  3. semantico-pragmatic factors such as semantic relations between RC and N1 and between N1 and N2 (e.g., Aoyama & Inoue, 2005)

Erratum in our paper text
Selected References


Katsika, Kalliopi (2011) Attachment preferences and corpus frequencies in PP ambiguities: Evidence from Greek. Selected Papers from the 19th ISTAL.

See our Proceedings paper for other references.

Thank you very much for your attention!

Comments welcome: toshiyamada@hotmail.co.jp