Hungarian relative clause processing: Diverging results in L-Maze and A-Maze

Wesley Orth
University of Toronto

Dávid Nemeskey
Eötvös Loránd University

Eszter Ronai
Northwestern University

37th Annual Conference on Human Sentence Processing
May 16-18, 2024
Relative clause processing asymmetry

Subject-extracted relative clause (SRC)

The engineer [who annoyed the analyst] wrote a report about the project.

Object-extracted relative clause (ORC)

The engineer [who the analyst annoyed] wrote a report about the project.

ORC is harder to process

(i.a. Gordon et al., 2001; Grodner & Gibson, 2005; King & Just, 1991; Ford, 1983; Traxler et al., 2002)
Two families of accounts

1. **Expectation**-based accounts (i.a., Hale, 2001; Levy, 2008)

2. **Memory**-based accounts (i.a., Gibson, 1998, 2000; Lewis & Vasishth, 2005)
Expectation-based account

Comprehenders predict upcoming structure based on previous experience.

The more expected a word is in its context, the easier it is to process:

$$\text{difficulty} \propto - \log P(w_i|w_{1:i-1}, \text{CONTEXT})$$

Estimated based on corpus frequencies.

English SRCs are more frequent than ORCs (i.a., Hale, 2001; Reali & Christiansen, 2007) → captures processing asymmetry.
Memory-based account

Syntactic structure is built incrementally
Dependency formation is sensitive to memory limitations

SRC: The engineer [who annoyed the analyst] wrote a report.

ORC: The engineer [who the analyst annoyed] wrote a report.

ORC requires a longer dependency → captures processing asymmetry

Cost: retrieval, storage, integration (Gibson, 1998, 2000)
similarity-based interference (Lewis & Vasishth, 2005)
**Hungarian: flexible word order**

<table>
<thead>
<tr>
<th>SRCs with short and long dependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>The engineer [who.NOM annoyed the analyst.ACC] ] reported the project-on</td>
</tr>
<tr>
<td>The engineer [who.NOM the analyst.ACC annoyed] ] reported the project-on</td>
</tr>
</tbody>
</table>

Both: ‘The engineer who annoyed the analyst wrote a report about the project.’
Hungarian: flexible word order

**ORCs with short and long dependencies**

A mérnök [akit idegesített az elemző ] beszámolt a projektről.
The engineer [who.acc annoyed the analyst.nom ] reported the project-on

A mérnök [akit az elemző idegesített ] beszámolt a projektről.
The engineer [who.acc the analyst.nom annoyed ] reported the project-on

Both: ‘The engineer who the analyst annoyed wrote a report about the project.’
Predictions for Hungarian RCs

Memory-based accounts:
longer dependencies harder
local < non-local

Location: RC Verb (place of verb-argument integration)
Predictions for Hungarian RCs

Expectation-based accounts:
less frequent structures harder

**non-local < local**
- global frequencies; conditional probabilities (Ronai & Xiang, 2023; Oravecz et al., 2014)
- anti-locality (i.a., Konieczny, 2000, 2005; Levy & Keller, 2013; Vasishth & Lewis, 2006)

Location: **RC Verb** (anti-locality, conditional probabilities)
Predictions for Hungarian RCs

Relative pronoun case marking signals RC structure (aki vs. akit)

Expectation-based accounts
SRCs more frequent than ORCs
SRC < ORC at Rel. Pronoun
Schematic summary of predictions

SRC

- local (shorter dependency)
  - less frequent

- non-local (longer dependency)
  - more frequent

ORC

- local (shorter dependency)
  - less frequent

- non-local (longer dependency)
  - more frequent
Prior work using self-paced reading

(Ronai & Xiang, 2023)

SPR subject to spillover (i.a., Roland et al., 2021)

Maze-task offers better effect localization (Boyce et al., 2020)
Method-based divergence in Russian RCs

**Russian RCs:** case-marking at relative pronoun; word order flexibility

**Relative pronoun:** SRC < ORC
- No clear effect found in SPR (Levy et al., 2013; Price & Witzel, 2017)
- **Confirmed in eye-tracking** while reading (Price & Witzel, 2017)

**RC verb:** local < non-local
- Locality effect in SPR (Levy et al., 2013; Price & Witzel, 2017)
- Anti-locality effect in eye-tracking (Price & Witzel, 2017)
# Experimental manipulation

## 2-by-2 design: RC type (SRC vs. ORC) x Verb Locality (local vs. non-local)

(Levy et al., 2013; Ronai & Xiang, 2023)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Subject</th>
<th>Rel. pr.</th>
<th>Pre-V NP</th>
<th>RC verb</th>
<th>Post-V NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC local</td>
<td>a mérnök</td>
<td>aki</td>
<td></td>
<td>idegesítette</td>
<td>az elemzőt</td>
</tr>
<tr>
<td>SRC non-local</td>
<td>a mérnök</td>
<td>aki</td>
<td>az elemzőt</td>
<td>idegesítette</td>
<td></td>
</tr>
<tr>
<td>ORC local</td>
<td>a mérnök</td>
<td>akit</td>
<td></td>
<td>idegesített</td>
<td>az elemző</td>
</tr>
<tr>
<td>ORC non-local</td>
<td>a mérnök</td>
<td>akit</td>
<td>az elemző</td>
<td>idegesített</td>
<td></td>
</tr>
</tbody>
</table>
L-Maze for Hungarian

Custom language extension for Wuggy (Python Version, Keuleers & Brysbaert 2010)

Hand corrections:

Regions & Morphology: Locality manipulation and morphology modification

SRC (local): I mulnád, epi reedenítálye éz elegült náp súl áze, bagyágort ö kroluktród.

ORC (non-local): I mulnád, epit éz elegül reedenítály náp súl áze, bagyágort ö kroluktród.
# Recap of predictions

<table>
<thead>
<tr>
<th>RC Verb Position</th>
<th>Rel. Pronoun</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Memory</strong></td>
<td><strong>Memory</strong></td>
</tr>
<tr>
<td>local &lt; non-local</td>
<td>No differences</td>
</tr>
<tr>
<td>Expectation</td>
<td>Expectation</td>
</tr>
<tr>
<td>non-local &lt; local</td>
<td>SRCs &lt; ORCs</td>
</tr>
</tbody>
</table>
L-Maze: results

L-Maze Response Time (N=46)

Response Time (ms)

Verb Locality
- Local
- Non-Local

RC Type
- ORC
- SRC

SRC p<0.001
Local p<0.05
Results from critical regions

L-Maze Response Time at Rel. Pronoun

SRC
p<0.001

Local
p<0.05

L-Maze Response Time at RC Verb

RC Type
ORC
SRC

Response Time (ms)

600 700 800 900 1000

Local Non-Local

Local Non-Local
A-Maze for Hungarian

A-Maze algorithm following Boyce, Futrell & Levy 2020

**Frequency:** Hungarian Webcorpus 2.0 (Nemeskey 2020)

**Language Model:** PULI-GPT-Trio (Yang, Laki, Váradi & Prószéky 2023)

Adjustments for transition from word-level LSTM to sub-word GPT model

Threshold adjustments for language specific frequency properties
**A-Maze for Hungarian**

**Hand Corrections**

**Regions & Morphology:** Locality manipulation and morphology modification

**Word Length:** Length matching issues with short words – foils hand selected

**Surprisal:** Matrix Noun and Matrix Verb checked for grammatical status

**SRC (local):** x-x-x körébe, női [tananyagának](#) jó rágózás apa még elő, készítünk ó kijavítást.

**ORC (non-local):** x-x-x körébe, nőit jó rágózó [tananyagnak](#) apa még elő, készítünk ó kijavítást.

<table>
<thead>
<tr>
<th>hogy</th>
<th>hasi</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>i</td>
</tr>
</tbody>
</table>
A-Maze: results

A-Maze Response Time (N=71)

Verb Locality
- Local
- Non-Local

RC Type
- ORC
- SRC

SRC p<0.001
Local P<0.001
Results from critical regions

A-Maze Response Time at Rel. Pronoun

SRC
p<0.001

A-Maze Response Time at RC Verb

SRC
p<0.001
Local
P<0.001
Theoretical implications

L-Maze Results:
- **SRC advantage** at relative pronoun (**Expectation**)
- **Locality** effect at RC verb (**Memory**)

Prior work has observed **multiple sources of processing cost**, often within the same structure
Russian RCs (Levy et al., 2013; Price & Witzel, 2017); German verb-final sentences (Levy & Keller, 2013, though cf. Vasishth et al. 2018); Hindi complex vs. simple predicates (Husain et al. 2014, though cf. Safavi et al. 2016 for Persian); German complex determiner (Schwab et al., 2022)

A-Maze Results:
- **SRC advantage** at relative pronoun (**Expectation**)
- **Anti-locality** effect at RC verb (**Expectation**)

Prior work has observed **different effects in different tasks** (Price & Witzel, 2017)
Methodological implications

Puzzle: locality vs. anti-locality at RC verb found in L-Maze vs. A-Maze:
- A-Maze forces incremental mode of processing $\rightarrow$ sharper expectations (Duff 2023)
- Engage with task as surprisal evaluation?

Maze effect detection better than SPR, in line with eye-tracking (relative pronoun)

Allows us to include relatively understudied languages in psycholinguistics
- Though challenges remain: word order flexibility non-trivial for A-Maze
Conclusion

Adaptation of Mazes to a language relatively understudied in psycholinguistics

Captured previously elusive expectation-based effect

Diverging findings based on L-Maze vs. A-Maze suggest need for further research into properties of the tasks
Thank you! (wes.orth@utoronto.ca; ronai@northwestern.edu)
Selected references

Price & Witzel. 2017. Sources of relative clause processing difficulty: evidence from Russian. JML.