An Analysis of Working Memory Constraints in a Head-Final Language

Evan Jaffe
Dept Linguistics, The Ohio State University
QP1

April 19, 2014
Question: Are working memory requirements similar cross-linguistically?

Method: Corpus analysis of center-embedding depth

Spoiler: Similar requirements for Japanese and English
Background - Sentence Processing

- Human sentence processing is subject to cognitive constraints
- Constraints should be cross-linguistic
- Constraints can inform models and make predictions
Background - Working Memory

- Chunking, Rehearsal, Memorization, Stimulus Mode
  - Magic Number 7 (+-2) [Miller, 1956]
  - Magic Number 4 (+-1) [Cowan, 2001]
  - Magic Number 2 [Gobet and Clarkson, 2004]
  - ...but not register overflow so much as interference [Lewis and Vasishth, 2005]
  - ...which elegantly gets degradation, confusability effects
Working Memory and Center-embedding

- (1) The rat [the cat [the dog chased] ate] died. [Chomsky and Miller, 1963]
- (2) You are what [what you eat] eats. [Pollan, 2006]

Zig-zag measure
Center-embedded sentences are difficult to process

- Imposes memory load on parser
  - [Johnson-Laird, 1983]
  - [Abney and Johnson, 1991]

Other explanations for processing difficulty

- Semantic, embedding categories [Karlsson, 2007]
- Distance, number of NPs [Gibson, 2000]
Previous Work

If embedding depth is related to a cognitive constraint, it should be cross-linguistically similar.

- Swedish - 4 connected components [Nivre, 2004]
- English - 3-4 connected components [Schuler et al., 2010]
- 7 Indo-European VO languages - 3 maximum depth in written [Karlsson, 2007]
- [Bader and Haussler 12] - corpus analysis of German, not strictly head-final
- dispute Karlsson’s competence/grammatical constraints and argue for performance limitations
- Japanese - ?
Why would SVO vs. SOV matter?

- Different embedding depths
- E.g., prenominal modification in object position

These trees are controversial! Alternatives will be discussed later.
Methods - Overview

- **Method**: corpus study of center-embedding depth
- Wall Street Journal section of Penn Treebank [Marcus et al., 1993]
- Switchboard Corpus [Godfrey et al., 1992]
- Kyoto University Corpus [Kawahara et al., 2002]
  - 40,000 sentences of newspaper text
  - automatically parsed, then hand-corrected
  - JUMAN Morphological Analyzer [Kurohashi and Nagao, 1998]
  - KNP Dependency Parser [Kurohashi and Nagao, 1994]
  - *bunsetsu* syntactic unit
    - One content word per *bunsetsu*
    - Particles and function words attached to nearest *bunsetsu* to the left
Bunsetsu examples

ロシア側は roshiagawaha
来てみると kitemiruto

(1) roshia-gawa-ha
russia-side-TOP
’the Russians’

(2) ki-te-mi-ru-to
come-TE-try-NONPAST-upon
’upon trying to come’

- Relative to English, longer ’words’, shorter sentences
- Articles, prepositions, some verbal auxiliaries
- Alternative tokenizations are possible
- Length normalization will be an issue
Methods - Depth

- Binarized trees
- Right-corner depth measure - left children of right parents increase depth
- Otherwise, children inherit depth of parent

\[
S_{L,1} \\
NP_{L,1} \quad IV_{R,1} \\
D_{L,1} \quad N_{R,1} \quad TV_{L,2} \quad NP_{R,1} \\
D_{L,2} \quad N_{R,1} \\
\text{the studio bought the publisher’s rights} \\
NP_{L,2} \quad G_{R,2} \\
D_{L,2} \quad N_{R,2} \\
\]

- greatest depth non-terminal defines maximum embedding depth
Methods - Convert dependency to category tree

Japanese dependency trees require conversion

- Bottom-up chart parsing style algorithm
- Finds increasingly larger continuous spans of satisfied dependencies
It is seen that Russia entered into the final discussion about control of the capital.
Results - Primary

- Consistent maximum depth as for English and Swedish

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>no connected components</td>
<td>26,200</td>
<td>28.68%</td>
<td>35</td>
<td>0.09%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>1 element</td>
<td>59,253</td>
<td>64.87%</td>
<td>3,101</td>
<td>8.14%</td>
<td>6,324</td>
<td>16.55%</td>
</tr>
<tr>
<td>2 elements</td>
<td>85,944</td>
<td>94.09%</td>
<td>23,536</td>
<td>61.76%</td>
<td>26,430</td>
<td>69.17%</td>
</tr>
<tr>
<td>3 elements</td>
<td>91,008</td>
<td>99.63%</td>
<td>36,433</td>
<td>95.61%</td>
<td>36,723</td>
<td>96.11%</td>
</tr>
<tr>
<td>4 elements</td>
<td>91,332</td>
<td>99.98%</td>
<td>38,039</td>
<td>99.82%</td>
<td>38,133</td>
<td>99.80%</td>
</tr>
<tr>
<td>5 elements</td>
<td>91,346</td>
<td>100.00%</td>
<td>38,105</td>
<td>99.99%</td>
<td>38,209</td>
<td>100.00%</td>
</tr>
<tr>
<td>6 elements</td>
<td>91,346</td>
<td>100.00%</td>
<td>38,107</td>
<td>100.00%</td>
<td>38,209</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

- Count and percent (cumulative) corpus coverage by maximum embedding depth for Switchboard, WSJ, Kyoto
Kyoto Max Embedding Depths by Length

An Analysis of Working Memory Constraints in a Head-Final Language
Kyoto Max Embedding Depths by Length (normed)

An Analysis of Working Memory Constraints in a Head-Final Language
WSJ Max Embedding Depths by Length
Switchboard Max Embedding Depths by Length
Switchboard Max Embedding Depths by Length (normed)

An Analysis of Working Memory Constraints in a Head-Final Language
Results - Secondary

- No significant effect of genre - WSJ vs. Switchboard
- Wilcoxon Rank Sum Test

<table>
<thead>
<tr>
<th>Data subset</th>
<th>Significance</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full corpora</td>
<td>***</td>
<td>2.2e-16</td>
</tr>
<tr>
<td>Fixed length bin (30-40)</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Top deciles</td>
<td>***</td>
<td>2.2e-16</td>
</tr>
</tbody>
</table>

- Length is a confound for predicting depth
- Top decile potentially still confounds depth - fixed bins trustworthy here
Results - Secondary

- Effect of language uncertain - WSJ vs. Kyoto
- Wilcoxon Rank Sum Test

<table>
<thead>
<tr>
<th>Data subset</th>
<th>Significance</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full corpora</td>
<td>***</td>
<td>2.2e-16</td>
</tr>
<tr>
<td>Fixed length bin (30-40)</td>
<td>***</td>
<td>2.2e-16</td>
</tr>
<tr>
<td>Top deciles</td>
<td>***</td>
<td>2.2e-16</td>
</tr>
</tbody>
</table>
Results - Secondary

- Likelihood Ratio Test - significant difference between top deciles
  - Model 1: depth ~length
  - Model 2: depth ~length + lang
  - Chisq: 420.64
  - Pr(>Chisq): 2.2e-16
- ...but top deciles still have different sentence length means (wsj=45.64, kyoto=28.48, swbd=25.07)
- ...length confound still potentially present
- Interaction term?
- How to reliably normalize sentence length cross-linguistically?
Conversion from dependency tree to category tree has multiple possibilities

[Bhatt and Xia, 2011]

[Xia and Palmer, 2001]
Alternate Left-branching binarization

An Analysis of Working Memory Constraints in a Head-Final Language
Left-branching binarization results

Kyoto Corpus Coverage by Tree Conversion Scheme

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>no connected components</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
<td>0.00%</td>
</tr>
<tr>
<td>1 element</td>
<td>4550</td>
<td>11.90%</td>
<td>6324</td>
<td>16.55%</td>
</tr>
<tr>
<td>2 elements</td>
<td>30,820</td>
<td>80.63%</td>
<td>26,430</td>
<td>69.17%</td>
</tr>
<tr>
<td>3 elements</td>
<td>37,589</td>
<td>98.14%</td>
<td>36,723</td>
<td>96.11%</td>
</tr>
<tr>
<td>4 elements</td>
<td>38,200</td>
<td>99.93%</td>
<td>38,133</td>
<td>99.80%</td>
</tr>
<tr>
<td>5 elements</td>
<td>38,225</td>
<td>100%</td>
<td>38,209</td>
<td>100%</td>
</tr>
</tbody>
</table>

Left-binarized Max Embedding Depth Mean: 1.881
Contiguous Max Embedding Depth Mean: 2.183
Implications

- Left-binarization predicts less working memory requirements
- Integration cost sites different predicts different places for slow down/speed up
- Compare model predictions to human subject data for reading times, ERP, etc.? [van Schijndel and Schuler, 2013]
- Spectrum of attachability
  - strict incremental attachment does not model memory
  - no attachment, memory demands spiral out of control
Possible...but might be less cognitively plausible
Contiguous model

- Constraint of only connecting components that share a head
- Must use additional memory once additional head is hypothesized
Future Work

- Corpus of Spoken Japanese - is the domain effect finding robust?
- Generate a latent-variable PCFG synthetic corpus - does it overgenerate center-embedding? [Schuler, 2011]
  - Train latent-variable PCFG parser on Kyoto
  - Generate corpus
  - Generate corpus with 5+ depth sentences removed
  - Compare corpus fit between each synthetic corpus and Kyoto
Conclusion

- Corpus analysis of a head-final language
- Maximum embedding depth is comparable to previous findings in head-initial languages
- No significant effect of genre (informal speech vs. written news)
- Uncertain effect of language
Thanks to my committee: Micha Elsner, Marie-Catherine de Marneffe, William Schuler
Thank you Mary Beckman, Cynthia Clopper, and Clippers members for additional feedback


