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

Evan Jaffe
Dept Linguistics, The Ohio State University
QP1

April 19, 2014

Introduction

- **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

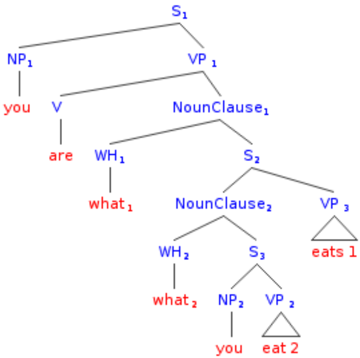
- Human sentence processing is subject to cognitive constraints
- Constraints should be cross-linguistic
- Constraints can inform models and make predictions

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

Working Memory and Center-embedding, cont.

- 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]

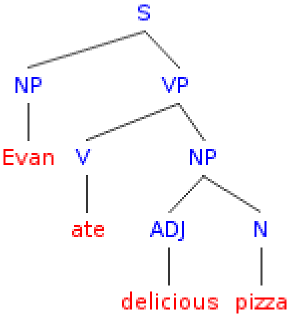
Previous Work

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

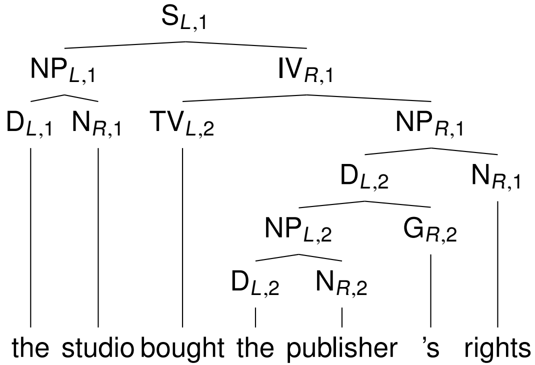
(1) roshia-gawa-ha
russia-side-TOP
'the Russians'

(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

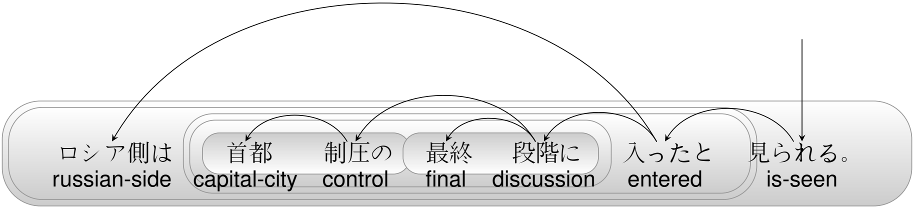
- Binarized trees
- Right-corner depth measure - left children of right parents increase depth
- Otherwise, children inherit depth of parent



- 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

Methods - Example Japanese Sentence

(3)	roshiagawaha	shuto	seiatsuno	saishuu	dankaini
	russia-side-TOP	capital	control-GEN	final	discussion-LOC

	haittato	mirareru
	enter-PAST	see-PASS

It is seen that Russia entered into the final discussion about control of the capital

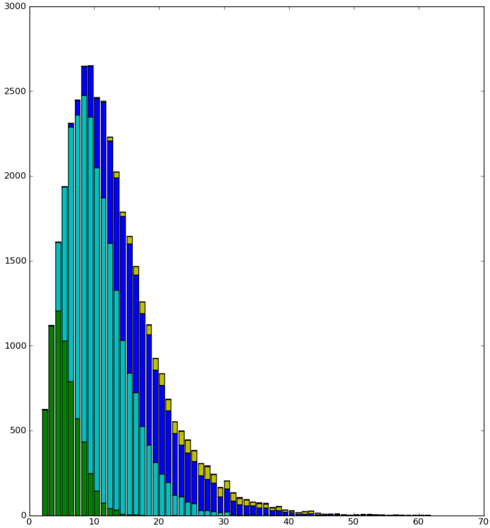
Results - Primary

- **Consistent maximum depth as for English and Swedish**

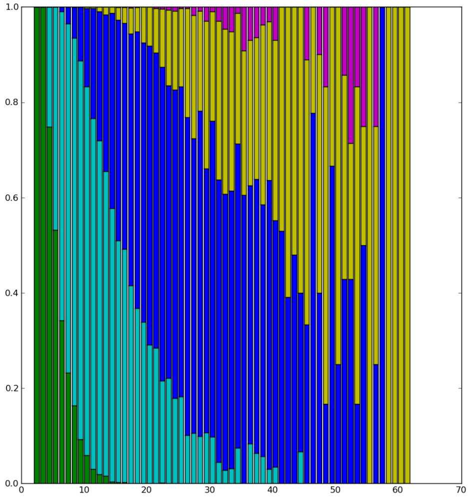
Memory capacity	Sent EN-swbd	Coverage EN-swbd	Sent EN-wsj	Coverage EN-wsj	Sent JP-kyo	Coverage JP-kyo
no connected components	26,200	28.68%	35	0.09%	0	0.00%
1 element	59,253	64.87%	3,101	8.14%	6,324	16.55%
2 elements	85,944	94.09%	23,536	61.76%	26,430	69.17%
3 elements	91,008	99.63%	36,433	95.61%	36,723	96.11%
4 elements	91,332	99.98%	38,039	99.82%	38,133	99.80%
5 elements	91,346	100.00%	38,105	99.99%	38,209	100.00%
6 elements	91,346	100.00%	38,107	100.00%	38,209	100.00%

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

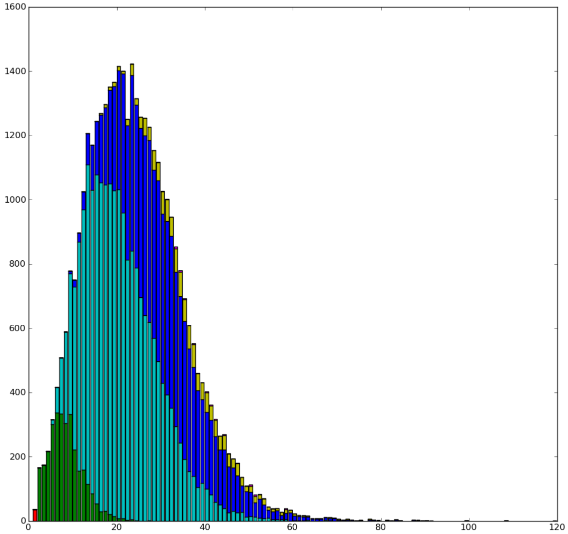
Kyoto Max Embedding Depths by Length



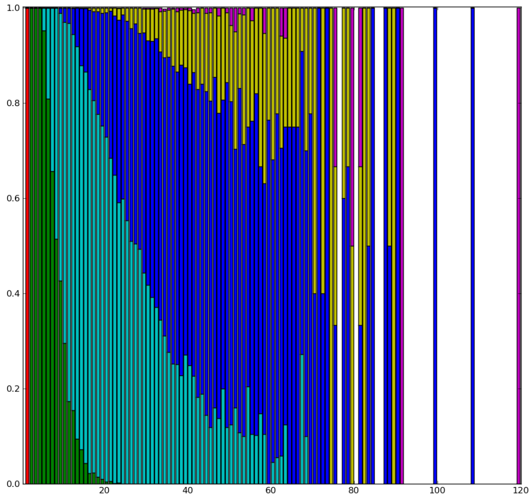
Kyoto Max Embedding Depths by Length (normed)



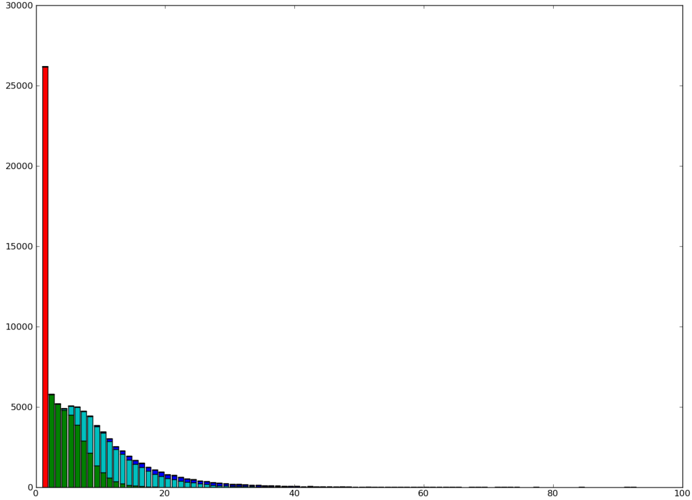
WSJ Max Embedding Depths by Length



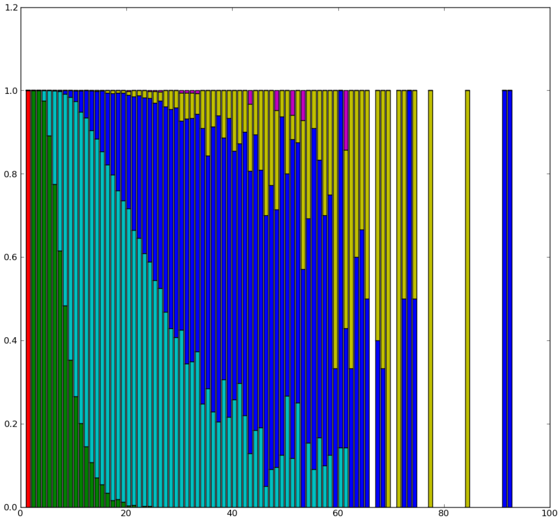
WSJ Max Embedding Depths by Length (normed)



Switchboard Max Embedding Depths by Length



Switchboard Max Embedding Depths by Length (normed)



Results - Secondary

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

Data subset	signficance	P-value
Full corpora	***	2.2e-16
Fixed length bin (30-40)		0.92
Top deciles	***	2.2e-16

- 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

Data subset	signficance	P-value
Full corpora	***	2.2e-16
Fixed length bin (30-40)	***	2.2e-16
Top deciles	***	2.2e-16

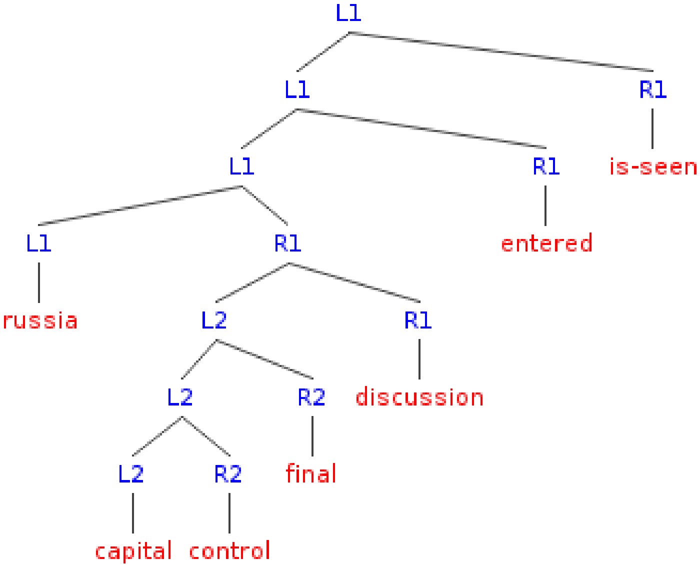
Results - Secondary

- 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?

Discussion

- Conversion from dependency tree to category tree has multiple possibilities
- [Bhatt and Xia, 2011]
- [Xia and Palmer, 2001]

Alternate Left-branching binarization



Left-branching binarization results

Kyoto Corpus Coverage by Tree Conversion Scheme

Memory capacity	Sent JP-left	Percent JP-left	Sent JP-cont	Percent JP-cont
no connected components	0	0.00%	0	0.00%
1 element	4550	11.90%	6324	16.55%
2 elements	30,820	80.63%	26,430	69.17%
3 elements	37,589	98.14%	36,723	96.11%
4 elements	38,200	99.93%	38,133	99.80%
5 elements	38,225	100%	38,209	100%

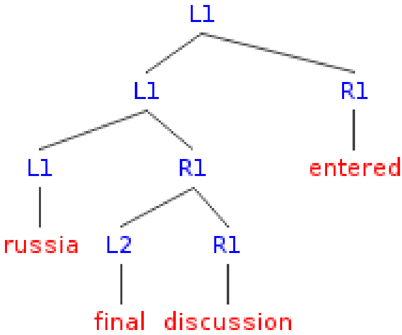
Left-binarized Max Embedding Depth Mean: 1.881

Contiguous Max Embedding Depth Mean: 2.183

Implications

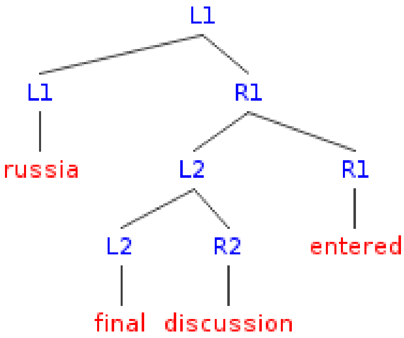
- Left-binarization predicts less working memory requirements
- Integration cost sites different predicts different places for slow down/speed up
- Compare model predictions to human subject data for reading times, ERP, etc.? [van Schijndel and Schuler, 2013]
- Spectrum of attachability
  - strict incremental 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!

Thanks to my committee: Micha Elsner, Marie-Catherine de Marneffe, William Schuler
Thank you Mary Beckman, Cynthia Clopper, and Clippers members for additional feedback

Bibliography I

Abney, S. P. and Johnson, M. (1991).
Memory requirements and local ambiguities of parsing strategies.
J. Psycholinguistic Research, 20(3):233–250.

Bhatt, R. and Xia, O. R. F. (2011).
Linguistic phenomena, analyses, and representations: Understanding conversion between treebanks.
In Proc. AFNLP.

Chomsky, N. and Miller, G. A. (1963).
Introduction to the formal analysis of natural languages.
In Handbook of Mathematical Psychology, pages 269–321. Wiley, New York, NY.

Bibliography II

Cowan, N. (2001).
The magical number 4 in short-term memory: A reconsideration of mental storage capacity.
Behavioral and Brain Sciences, 24:87–185.

Gibson, E. (2000).
The dependency locality theory: A distance-based theory of linguistic complexity.
In Image, language, brain: Papers from the first mind articulation project symposium, pages 95–126, Cambridge, MA. MIT Press.

Gobet, F. and Clarkson, G. (2004).
Chunks in expert memory: Evidence for the nagical number four...or is it two?
Memory, 6(12).

Bibliography III

Godfrey, J. J., Holliman, E. C., and McDaniel, J. (1992).
Switchboard: Telephone speech corpus for research and development.
In Proc. ICASSP, pages 517–520, San Francisco, CA.

Johnson-Laird, P. N. (1983).
Mental models: towards a cognitive science of language, inference, and consciousness.
Harvard University Press, Cambridge, MA, USA.

Karlsson, F. (2007).
Constraints on multiple center-embedding of clauses.
Journal of Linguistics, 43:365–392.

Bibliography IV

Kawahara, D., Kurohashi, S., and Hasida, K. (2002).
Construction of a Japanese relevance-tagged corpus (in Japanese).
In Proceedings of the Eighth Annual Meeting of the Association for Natural Language Processing, pages 495–498. Association for Natural Language Processing.

Kurohashi, S. and Nagao, M. (1994).
A syntactic analysis method of long japanese sentences based on the detection of conjunctive structures.
In Proc. ACL.

Kurohashi, S. and Nagao, M. (1998).
Japanese Morphological Analysis System JUMAN version 3.6.

Bibliography V

Lewis, R. L. and Vasishth, S. (2005).
An activation-based model of sentence processing as skilled memory retrieval.
Cognitive Science, 29(3):375–419.

Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1993).
Building a large annotated corpus of English: the Penn Treebank.
Computational Linguistics, 19(2):313–330.

Miller, G. A. (1956).
The magical number seven, plus or minus two: Some limits on our capacity for processing information.
Psychological Review, 63:81–97.

Bibliography VI

Nivre, J. (2004).
Incrementality in deterministic dependency parsing.
In Proceedings of the Workshop on Incremental Parsing: Bringing Engineering and Cognition Together (IncrementParsing '04), pages 50–57.

Pollan, M. (2006).
In Defense of Food: An Eater's Manifesto.
Penguin Books, New York.

Schuler, W. (2011).
Effects of filler-gap dependencies on working memory requirements for parsing.
In Proceedings of COGSCI, pages 501–506, Austin, TX. Cognitive Science Society.

Bibliography VII

- Schuler, W., AbdelRahman, S., Miller, T., and Schwartz, L. (2010).
  Broad-coverage incremental parsing using human-like memory constraints.
  Computational Linguistics, 36(1):1–30.
- van Schijndel, M. and Schuler, W. (2013).
  An analysis of frequency- and recency-based processing costs.
  In Proceedings of NAACL-HLT 2013. Association for Computational Linguistics.
- Xia, F. and Palmer, M. (2001).
  Converting dependency structures to phrase structures.
  In Proc. HLT.