An Analysis of Frequency- and Memory-Based Processing Costs

Marten van Schijndel and William Schuler

Linguistics, The Ohio State University

Introduction

Frequency has a huge effect on processing speeds, but what causes frequency? Perhaps cognitive limitations (e.g. in memory) lead to preferred structures. How do current theories of working memory fit with current theories of language processing? What influence do memory limitations have on reading times?

Frequency Effects

Surprisal

How unpredictable a word is given the preceding context [Hale, 2001]:

$$surprisal(x_t) = -\log_2 \left(\frac{\sum_{s \in S(x_1 \ldots x_t)} P(s)}{\sum_{s \in S(x_1 \ldots x_{t-1})} P(s)} \right) \quad (1)$$

where $S(x_1 \ldots x_t)$ is the set of syntactic trees whose leaves have $x_1 \ldots x_t$ as a prefix.

Entropy Reduction

Entropy is a measure of uncertainty:

$$H(x_{1\ldots t}) = \sum_{s \in S(x_1 \ldots x_t)} -P(s) \cdot \log_2 P(s) \quad (2)$$

The reduction in uncertainty caused by observing x_t [Hale, 2003]:

$$\Delta H(x_{1\ldots t}) = \max(0, H(x_{1\ldots t-1}) - H(x_{1\ldots t})) \quad (3)$$

References

[Botvinick, 2007] Botvinick, M. (2007). Multilevel structure in behavior and in the brain: a computational model of Fuster's hierarchy. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, 362:1615–1626.

[Gibson, 2000] 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.

[Hale, 2001] Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of NAACL*, pages 159–166.

[Hale, 2003] Hale, J. (2003). *Grammar, Uncertainty and Sentence Processing*. PhD thesis, Cognitive Science, Johns Hopkins University.

[Just and Varma, 2007] Just, M. A. and Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, & Behavioral Neuroscience*, 7:153–191.

[Kennedy et al., 2003] Kennedy, A., Pynte, J., and Hill, R. (2003). The Dundee corpus. In *Proceedings of ECEM 12*.

[Lewis et al., 2006] Lewis, R. L., Vasishth, S., and Dyke, J. A. V. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, 10(10):447–454.

[van Schijndel et al., 2012] van Schijndel, M., Exley, A., and Schuler, W. (2012). Connectionist-inspired incremental PCFG parsing. In *Proceedings of CMCL-NAACL 2012*.

[Wu et al., 2010] Wu, S., Bachrach, A., Cardenas, C., and Schuler, W. (2010). Complexity metrics in an incremental right-corner parser. In *Proceedings of ACL*, pages 1189–1198.

Memory and Processing

Connected Components

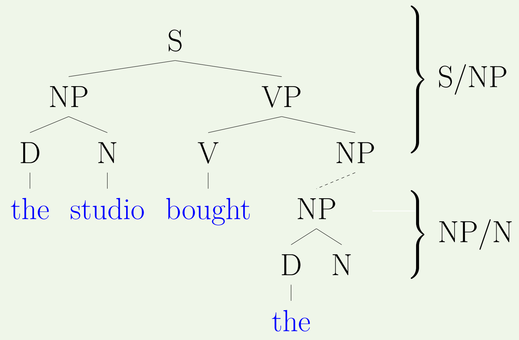
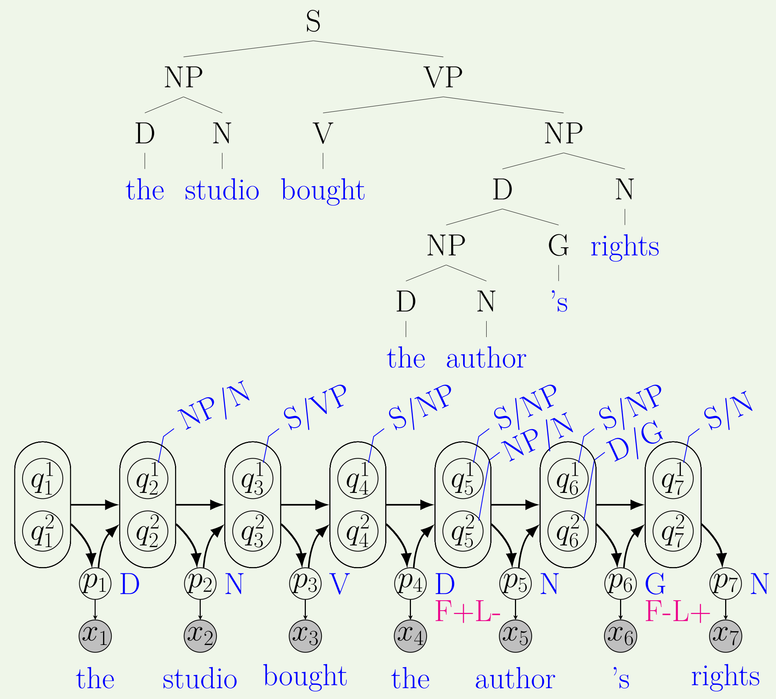


Figure 1: Two disjoint connected components of a phrase structure tree for the sentence *The studio bought the author's rights*, shown immediately prior to the word *author*.

Parsing Operations:
F = Word is the First element of a connected component
L = Word is the Last element of a connected component



Memory Theories

Dependency Locality Theory

[Gibson, 2000]

Difficulty { Unresolved dependencies; Memory operations

Storage cost { Beginning dependencies; Maintaining dependencies

Integration cost { Resolving dependencies

ACT-R

[Lewis et al., 2006]

Difficulty { Activation decay; Similarity interference

Encoding cost { Beginning a new dependency

Retrieval cost { Resolving a dependency

Retrieval can be *facilitated* by re-activations.

Hierarchic Sequential Prediction

[Botvinick, 2007]

- Learned *sequential* associations
- Contextual *temporal* associations

Difficulty { Temporal cueing

Temporal cueing { Resolving embedded dependencies

Dynamic Recruitment

[Just and Varma, 2007]
Difficult constructions → extra processing resources

Difficulty { Center embeddings

Recruitment { Beginning embeddings

Release { Completing embeddings

Embedding Difference

[Wu et al., 2010]

Difficulty { Dependency maintenance

- **Increased embedding depth**
- **Reduced embedding depth**

Predictions

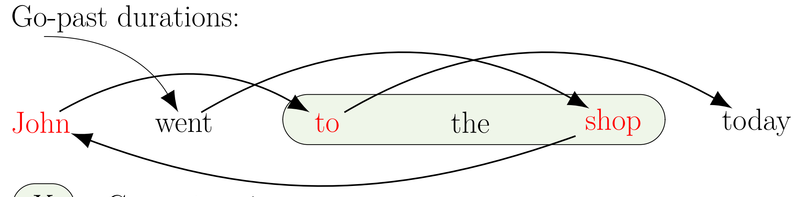
Theory	F+L–	F–L+
Dependency Locality Theory	**positive**	**positive**
ACT-R	**positive**	**positive**
Hier. Sequential Prediction		**positive**
Dynamic Recruitment	**positive**	**negative**
Embedding Difference	**positive**	**negative**

Table 1: Predicted correlation between each operation and reading times under each theory of working memory. Hierarchic sequential prediction is agnostic about the processing speed of F+L– operations.

Eye Tracking

- Slower reading = difficulty
- How much can be processed up to a given point?

Go-past durations:



X = Go-past region
Red = Fixation in go-past duration

Training

Parser and Lexicon: WSJ 02-21
- 950,028 words
- 39,832 sentences
- Accuracy = Berkeley Parser [van Schijndel et al., 2012]

N-grams: Brown, WSJ 02-21, BNC, Dundee
- 87,302,312 words
- 5,052,904 sentences
- Smoothed with modified Kneser-Ney

Evaluation

Dundee Eye-tracking Corpus [Kennedy et al., 2003]
- 10 subjects
- 2,388 sentences
- 154,168 go-past durations

Omit:
- Unknown words (< 5 times in WSJ)
- First and last of each line
- Fixations after long saccades (> 4 words)

Mixed Effects Model:
- Word length
- Prev/Next fixated?
- Unigram and bigram
- Sentence position
- Length of region
- Subject/Item random intercepts
- Cum. surprisal
- Surprisal
- Cum. entropy reduction
- Joint interactions
- Factors from prev. region

Non-cumulative metrics calculated from initial word of region
Cumulative metrics summed over region

Factor	Operation	Coeff	t-score	p-value
F–L–	Cue Active	0.0026	0.60	0.55
F+L–	**Initiate**	**0.0224**	**7.10**	**$2.22 \cdot 10^{-14}$**
F–L+	**Integrate**	**-0.0154**	**-5.44**	**$5.23 \cdot 10^{-8}$**
F+L+	Cue Awaited	-0.0031	-1.55	0.12

Conclusions

Support for dynamic recruitment:
- Initiation difficulty due to load causes recruitment
- Returns processing to average difficulty
- Integration facilitation due to reduced load causes release
- Returns processing to average difficulty

Initiation inhibition suggests difficulty beyond frequency effects, which perhaps causes infrequency of embeddings.