Modeling sentence processing in ACT-R

Shravan Vasishth

Department of Computational Linguistics
Saarland University, PO Box 15 11 05
66041 Saarbrücken, Germany
vasishth@acm.org

Richard L. Lewis

Department of Psychology
University of Michigan
Ann Arbor, MI, USA
rickl@umich.edu

Abstract

We present a series of simulations of behavioral data by casting a simple parsing model in the cognitive architecture ACT-R. We show that constraints defined in ACT-R, specifically, those relating to activation, can account for a range of facts about human sentence processing. In doing so, we argue that resource limitation in working memory is better defined as an artefact of very general constraints on information processing.

1 Introduction

It is clear that human sentence parsing is affected by the cognitive system's constraints: parsing does not bypass the brain and therefore must be affected by its information processing capabilities.

Although this point has been addressed extensively in the connectionist literature, we present a somewhat different approach to this problem by casting parsing within the cognitive architecture, ACT-R (Anderson et al., 2002) and directly using the constraints provided in ACT-R to account for several interesting cross-linguistic facts: the well-known sentential complement/relative clause asymmetry and the subject/object relative clause asymmetry in English (Gibson, 2000), (Grodner and Gibson, 2003); some recent results (Vasishth, 2003) involving Hindi center embeddings, including a principled account of individual variation in subject behavior; and a surprising result

where more retrieval cues actually hinder processing (Agnihotri et al., 2003).

In developing this approach, we argue that resource limitation in working memory is better defined as an artefact of very general constraints on information processing – specifically, rehearsal and spreading activation – rather than as an inherent numerical bound on memory capacity (cf. (Gibson, 2000; Hawkins, 1994); also see Section 3.6).

In the rest of this paper, we first introduce the ACT-R architecture. Then we present the results of several simulations of experiments available in the psycholinguistic literature. The paper concludes with a discussion of the potential advantages and shortcomings of this approach, and of the broader consequences of modeling parsing within a cognitive architecture.

2 A brief introduction to the cognitive architecture ACT-R

ACT-R is a theory about human cognition. It is implemented as a programming environment within common lisp, and allows the development of computational models that can closely simulate experimental methodologies such as eye-tracking and self-paced reading. ACT-R has been used to model a wide array of behavioral data from learning and memory, problem solving and decision making, language and communication, perception and attention, cognitive development, and individual differences (Anderson et al., 2002).

The ACT-R architecture is attractive as a modeling tool for three reasons. First, it is based on

a wide array of empirical results in various domains of cognitive psychology. Second, it is flexible enough to permit the modeler to add their own assumptions and theories about the specific task to be modeled. Finally, ACT-R models yield dependent measures such as reading time in much the same way as humans performing the experiment; e.g., the system can easily be programmed to simulate key presses after it processes material presented on the screen.
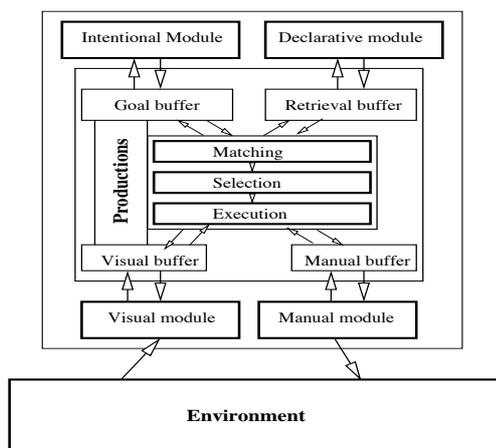

Figure 1: This is a schematic view of the ACT-R system. "Environment" is the outside world that ACT-R is programmed to interact with. The arrows show the possible flows of information. Productions and the central box with the boxes labeled "Matching", "Selection", and "Execution" are intended to represent a set of central executive mechanisms and processes.

As shown in Figure 1, the architecture consists of several MODULES such as Declarative, Visual, and Manual. Each module is associated with a BUFFER which temporarily stores information for a given action. For example, the visual buffer is used to store an item "seen" by the system in the environment before it is used in the service of some task.

The module that is especially important for the present paper is the Declarative (henceforth, DM). DM represents permanent memory: every fact that is assumed to be known is encoded as a CHUNK in declarative memory. A chunk is an attribute-value

list structure with a special attribute, ISA, which defines its type. The attributes are also referred to as slots. The value of a chunk's slot is also (by definition) a chunk, unless it is double-quoted or is the lisp primitive "nil".

Each DM chunk has an activation that determines its speed of retrieval, and the probability that it will be retrieved; the initial activation for a given chunk can be set manually.

There is a GOAL BUFFER that holds a current goal under consideration (there can be only one goal at one time); this goal is a chunk with a given type and possibly instantiated slots.

The control structure for modeling a sequence of events is a set of PRODUCTIONS; a production is simply an if-then statement of the following general form: for a given state of one or more buffers and/or DM, execute some actions. Examples of executing actions are retrieving something from DM; changing a value in one of the goal's slots; repositioning the hand over a keyboard; a visual shift of attention; changing the goal to a new one, etc. If the goal is changed, then this new goal now occupies the goal buffer.

Building an ACT-R model is essentially a definition of possible sequences of actions for a given state of affairs. Events like retrievals from DM are triggered by looking at the contents of one or more buffers. For example, the ACT-R system "sees" an item/object on the screen and then encodes it as a visual chunk. This chunk can then be harvested from the visual buffer; it includes (as slot-value specifications) information about the content of the item seen, its x-y coordinates, etc. One can define an action based on this information, such as retrieving a chunk from DM.

3 Modeling sentence parsing in ACT-R

Previous research suggests that humans employ some variant of left-corner (LC) parsing (see, e.g. (Resnik, 1992)) In its essence, LC parsing involves a bottom-up and a top-down (predictive) step.

In the simulations we present below, we assume that sentence structure templates are available in declarative memory as underspecified chunks. We illustrate the parsing process with a simple example (Figure 2). Suppose that the sentence to be parsed is *The girl ran*, and suppose that we are

simulating self-paced reading (Just et al., 1982). When the word *the* is seen, a bottom-up and top-down structure building step results in a sentence with an intransitive verb being predicted. This structure becomes the current goal. Then the word *girl* is seen and processed, i.e., its lexical entry is retrieved from declarative memory. The noun slot in the goal is then instantiated with that lexical entry. In the next step, if the word *ran* is seen the relevant lexical item for the verb is retrieved and instantiated with the verb slot of the goal; here, the verb's argument is also retrieved and integrated with the subcategorization frame of the verb. If, instead of *ran* the word *that* appears, a new goal is created, with any previously instantiated slots of the preceding goal being passed on to the new goal, and parsing proceeds from there.
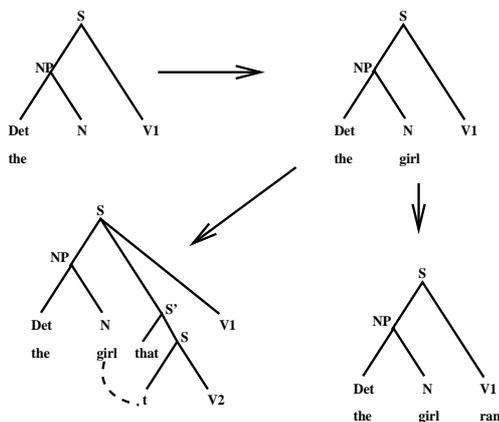

Figure 2: A simple illustration of parsing steps in the ACT-R simulations presented.

Each retrieval of an item from memory results in a surge in its activation, and the higher the activation of an item the faster it is processed. At the same time, activation decays according to the power law of forgetting (Anderson et al., 2002). These and other ACT-R mechanisms are discussed in more detail in the next section.

In our simulations, for simplicity we code in the exact steps that ACT-R takes for particular sentences. Although it is feasible to build a very general parser in pure ACT-R, before doing this we wanted to first establish whether ACT-R's reactivation mechanisms can account for a reasonable array of facts from the sentence processing literature. In forthcoming work (to appear in a special

issue of Cognitive Science) we provide a detailed description of a model employing essentially identical mechanisms described here, but one that behaves more like a standard parser.

The next section presents some relevant modeling results.

3.1 English subject versus object relative clauses

It is well known (Grodner and Gibson, 2003) that English subject relatives are easier to process that object relatives (1). In the parsing model outlined above, we can model this result without changing any ACT-R parameters at all (i.e., we use the default settings for the parameters).

(1)   a. The reporter who sent the photographer to the editor hoped for a good story.

      b. The reporter who the photographer sent to the editor hoped for a good story.
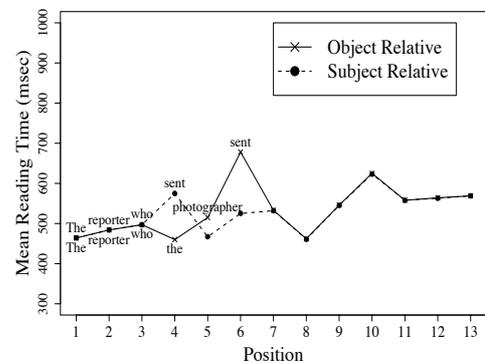

Figure 3: The reading times provided by the model. Retrieval of *reporter* at *sent* is harder in the object relative because of increased argument decay.

The explanation comes from the decay of the arguments of the verb *sent*: in object relatives the argument *reporter* decays much more than in the subject relative by the time it is integrated with the verb's subcategorization frame (Figure 3). This is because more time elapses between the argument

being first seen and its retrieval at the verb.[1]

3.2 The SC/RC asymmetry in English

It is also well-known (Gibson, 2000) that a sentential complement (equivalently, complement clause) followed by a relative clause is easier to process than the reverse ordering:

(2) a. The fact that the employee who the manager hired stole office supplies worried the executive.

b. #The executive who the fact that the employee stole office supplies worried hired the manager.

RC/SC (hard); total RT = 7605 msec
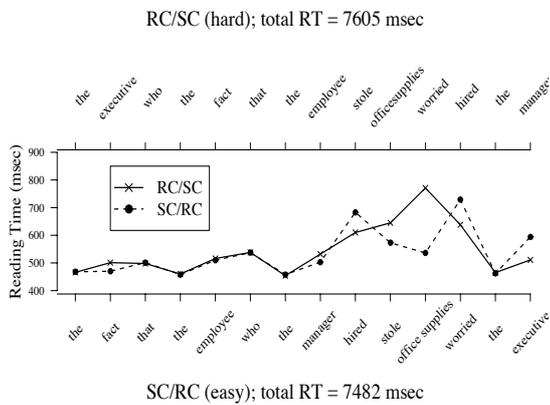

SC/RC (easy); total RT = 7482 msec

Figure 4: Model's behavior in the complement-clause/relative-clause contrast.

As in the previous discussion about relative clauses, in the harder case the decay of the argument *executive* at the verb *worried* is greater compared to the decay of the argument *employee* at *hired* in the easier-to-process sentence. In addi-

---

tion, the total reading time for the harder sentence is about 120 msec longer.[2]

3.3 Hindi center embeddings

Previous work (Hakes, 1972), (Konieczny, 2000) has shown that if argument-verb distance is increased, processing is easier at the verb. (Vasishth, 2003) presented similar results in Hindi. The Hindi experiment manipulated distance by comparing the baseline condition (3a) with the case where an adverb intervened (3b), a verb-modifying PP intervened (3c), and relative clause intervened that modified the preceding NP (3d).

(3) a. Siitaa-ne Hari-ko Ravi-ko
       Sita-erg Hari-dat Ravi-dat
       [kitaab-ko khariid-neko] bol-neko
       book-acc buy-inf   tell-inf
       kahaa
       told
       'Sita told Hari to tell Ravi to buy the book.'

b. Siitaa-ne Hari-ko Ravi-ko
   Sita-erg Hari-dat Ravi-dat
   [kitaab-ko **jitnii-jaldii-ho-sake**
   book-acc   as-soon-as-possible
   khariid-neko] bol-neko kahaa
   buy-inf   tell-inf told
   'Sita told Hari to tell Ravi to buy the book as soon as possible.'

c. Siitaa-ne Hari-ko Ravi-ko
   Sita-erg Hari-dat Ravi-dat
   [kitaab-ko **ek baṛhiya dukaan se**
   book-acc   from-a-good-shop
   khariid-neko] bol-neko kahaa
   buy-inf   tell-inf told
   'Sita told Hari to tell Ravi to buy the book from a good shop.'

d. Siitaa-ne Hari-ko Ravi-ko
   Sita-erg Hari-dat Ravi-dat

[kitaab-ko **jo-mez-par-thii**
book-acc that-was-on-a-table
khariid-neko] bol-neko kahaa
buy-inf tell-inf told
'Sita told Hari to tell Ravi to buy the
book that was lying on a/the table.'

In all the "insertion" cases a statistically significant speedup was observed at the verb, compared to the baseline condition.

This experiment was modeled in the ACT-R system by using our assumption that the goal (predicted syntactic structure) is reactivated each time it (i.e., the entire predicted structure) is modified. The intervening items result in an extra retrieval compared to the baseline, resulting in faster processing at the verb. In this model, one parameter was changed: the rate of decay of items. We justify this change in the next sub-section.
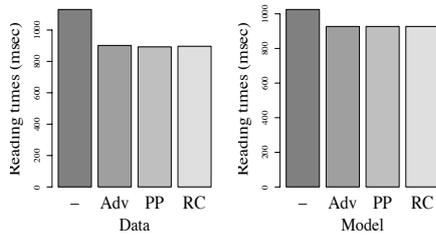
The modeling results are shown in Figure 5.

Figure 5: Reading times from data versus model, at the first verb.

3.4 Individual variation in Hindi center embedding data

In the Hindi experiment, there was a further variation in the data when individual subjects' data were considered: only about 48% of subjects showed a speedup at the verb. About 21% showed a slowdown and there was only a few milliseconds difference (essentially no difference) in the reading times for about 31% of the subjects. The observed variation was a systematic trend in the sense that the 47% of the subjects who showed a speedup or slowdown in adverb-insertion case also showed the *same* trend in the PP- and RC-inserted

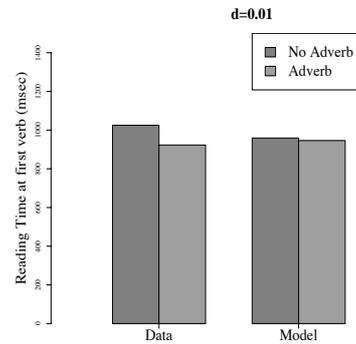cases – the probability of this happening is considerably below chance level.
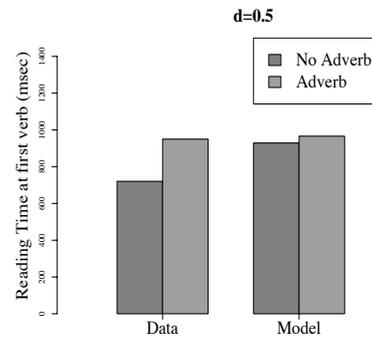

Figure 6: Modeling speedup.
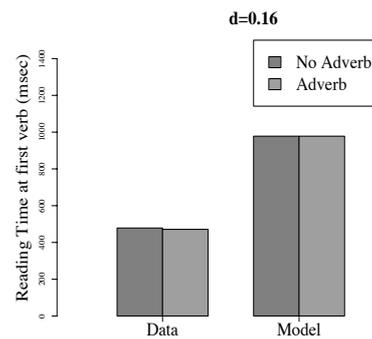

Figure 7: Modeling slowdown.

Figure 8: Modeling no difference in reading time.

The rate of decay defined in ACT-R's rehearsal equation can systematically explain this variation. Consider the situation where a chunk i with an initial activation of B_i is retrieved. The activation is

recalculated each time a retrieval occurs, according to the following equation.

$$(4) \quad B_i = \ln(\sum_{j=1}^{n} t_j^{-d})$$

Here, n is the number of times the chunk i was successfully retrieved, t_j is the time elapsed since the j-th retrieval, and d is a decay rate that defaults to 0.5 in ACT-R. This equation reflects the log odds that a chunk would reoccur as function of how it has appeared in the past (Anderson et al., 2002, 17).

It turns out that the d parameter take us beyond boolean predictions: $d = .01$ results in a speedup; $d = .5$ results in a slowdown; and $d = .16$ results in no difference in RT at the verb; see Figures 6 to 8.[3]

3.5 Modeling the effect of agreement cues

Recent work (Agnihotri et al., 2003) suggests that increasing the number of retrieval cues for an argument at a verb from one to two can result in *increased* processing difficulty.

The data can be summarized as follows. In Hindi, when both the subject and object NP are case marked, the verb has no agreement morphology (hereafter, No Agreement, NA). By contrast, when one NP has no overt case morphology, the verb agrees with this bare NP and shows the relevant agreement morphology (object agreement, OA). To illustrate, (5a) shows the NA situation, and (5b) shows OA.

(5) a. baccii-ko subah   jaldi-se laṛke-ne
         girl-acc   morning quickly boy-erg
         bulaayaa
         called
         'The boy called the girl quickly in the morning.'

     b. **baccii** subah   jaldi-se laṛke-ne
         girl   morning quickly boy-erg
         bulaay**ii**
         called

'The boy called the girl quickly in the morning.'

This was a self-paced reading study and showed that when agreement morphology is present on the verb, there is a longer reading time (i.e., increased processing difficulty) at the verb. This result can be modeled within the ACT-R architecture using its definition of spreading activation.

In ACT-R the values of the slots of a goal in a given production spread activation to matching chunks in DM, and this increases activation according to the following equation.

$$(6) \quad A_i = B_i + \sum_{k=1}^{m}(P_k \times M_{ki}) + \sum_{j=1}^{n}(W_j \times S_{ji})$$

The equation says: for a given chunk i, the activation will depend on its base-level activation, plus any mismatch penalty[4], plus ACTIVATION SPREAD from the n sources of activation in the goal.
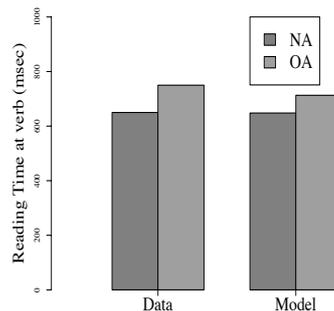

Figure 9: Modeling results of the Hindi agreement experiment

Focusing on spreading activation alone, suppose the system is examining a chunk i as a possible candidate for retrieval. Suppose j slots and their values in the goal are identical with slots and their values in chunk i. Each of the j slots' values of the current goal is assumed to have a weight W_j (this represents the attentional weighting of the slots that are part of the current goal) and S_{ij} is the strength of association from slot value j to

---

[3]Of course, modeling individual variation in terms of differing rates of decay assumes that subjects exhibit varying degrees of decay rates. An experiment is currently in progress that attempts to correlate varying verbal sentence span with subject behavior in the insertion cases.

[4]We do not address this part of the equation in this paper and keep partial matching disabled in our simulations; however, partial matching can be used to model several key results in inhibition effects during lexical access.

chunk i. W_j is determined by $W/n = W_j$, where n is the number of slots involved in the goal; W is a free parameter in ACT-R, with a default value 1.

S_{ij} is determined by the following equation; here, S is also a free parameter, called maximum associative strength.

(7) $S_{ij} = S - \ln(\text{fan}_i)$

The term "fan" here means: for a given source slot value S1, the number of chunks in DM that have a matching (identical) slot value in their own feature (slot value) specification.

We can model the experiment result using these equations by adding another source of activation in the current goal – the gender marking – when the verb is being retrieved, and by setting the free parameter S to -3. The fit of the model with the data are shown in Figure 9.

Note that it is equally possible to model the opposite situation as well, by setting S to $+3$. Only further experiments can establish whether in sentence parsing -3 falls within a bounded range of values for the S parameter. The point here is simply that ACT-R's domain-independent activation equation can explain a domain-specific phenomenon.

3.6   Comparison with other models

The model presented here is very different in conception from existing models of sentence processing. For example, consider Early Immediate Consituents (Hawkins, 1994) and Discourse Locality Theory (Gibson, 2000), two theories with significant empirical coverage. Both theories propose variants of what we will call the distance hypothesis: increasing the distance between arguments and a subsequently appearing verb (head) that selects for them results in increased processing difficulty at the verb. Distance here is quantified in terms of the number of words in a constituent (EIC) or the number of new discourse referents introduced between the arguments and head (DLT).

The present model claims that distance effects are actually a result of argument decay. Evidence that it is really decay and not EIC- or DLT-defined distance comes from a recent self-paced listening experiment (Vasishth et al., 2004) in which two

conditions were contrasted: arguments and verbs with (a) an adjunct intervening, (b) silence:

(8)   a.   vo-kaagaz / jisko   us-laṛke-ne /
           that-paper   which that-boy-erg
           mez ke-piiche gire-hue / dekhaa /
           table behind   fallen   saw
           bahut-puraanaa thaa
           very-old   was
           'That paper which that boy saw fallen
           behind a/the table was very old.'

      b.   vo-kaagaz / jisko   us-laṛke-ne /
           that-paper   which that-boy-erg
           SILENCE / dekhaa / bahut-puraanaa
                         saw   very-old
           thaa
           was
           'That paper which that boy saw was
           very old.'

In (8), the arguments *kaagaz*, 'paper', and *laṛkaa*, 'boy' are separated from the verb *dekhaa*, 'saw' by an adjunct containing two[5] discourse referents (8a); or by silence (8b). Subjects were allowed to interrupt the silence and continue listening to the rest of the sentence whenever they wanted to. Subjects interruped the silence (on an average) after about 1.4 seconds.

Distance based theories predict that having an intervening adjunct that introduces discourse referents should result in greater processing difficulty at the verb *dekhaa*, 'saw', compared to when silence intervenes. If decay rather than distance is the critical factor here that affects processing, then there should be greater difficulty at the verb in the silence condition than when in the items intervene (see Section 3.3 for why intervening items may facilitate processing). The results support the activation account: introducing silence results in significantly longer reading times at the verb *dekhaa* than when intervening items occur.

4   Conclusion

These preliminary modeling efforts suggest that very general constraints on information processing

---

[5]In DLT finite verbs also assumed to introduce a discourse referent.

can provide a principled account of parsing phenomena, and also brings human sentence parsing in line with the mainstream view in cognitive psychology (Miyake and Shah, 1999) that constraints on human working memory are likely to be an artefact of other cognitive constraints, such as activation.

That said, there are of course certain potential limitations in the work presented here. One problem is that several alternative hypotheses remain to be explored: the role of competence grammar and its own (possibly theory-internal) operations on processing; the role of experience (Crocker and Brants, 2000), and whether alternative cognitive architectures (Lewis and Young, 1999) or other theories of memory (Van Dyke, 2002) can provide a better account. However, the present research is a necessary first step since it provides a basis for such a comparison.

Secondly, there are specific assumptions in the modeling results that may be controversial. For example, we assume that entire sentence structures are predicted as goal chunks, and not verb-types (cf. (Konieczny, 2000)). We are conducting further experiments to explore the predictions made by different assumptions.

Finally, we have used toy simulations to explore the ACT-R constraint-interaction space, the task of scaling up such a model to parse essentially any kind of input is still in the future. However, we believe that the results presented are suggestive of the way in which a cognitively-oriented parser could be constructed.

Acknowledgements

References

Rama Kant Agnihotri, Anoop Mahajan, Shravan Vasishth, Gisbert Fanselow, and Matthias Schlesewsky. 2003. More isn't always better: A surprising constraint on retrieval cues in human sentence parsing. In *Proceedings of the AMLaP conference*, Glasgow, Scotland.

J.R. Anderson, D. Bothell, M.D. Byrne, and C. Lebiere. 2002. An integrated theory of the mind. MS, available from http://www.act-r.psy.cmu.edu/papers/403/IntegratedTheory.pdf.

M. W. Crocker and T. Brants. 2000. Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research*, 29(6):647–669.

Edward Gibson. 2000. Dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita, and Wayne O'Neil, editors, *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*. MIT Press, Cambridge, MA.

Daniel Grodner and Edward Gibson. 2003. Consequences of the serial nature of linguistic input. MS.

David T. Hakes. 1972. On understanding sentences: In search of a theory of sentence comprehension. Microfilm, University of Texas, Austin.

John A. Hawkins. 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press, New York.

M. A. Just, P. A. Carpenter, and J. D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111(2):228–238.

Lars Konieczny. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6):627–645.

R. L. Lewis and R. M. Young. 1999. The Soar cognitive architecture and human working memory. In Akira Miyake and Priti Shah, editors, *Models of working memory*. Cambridge University Press, Cambridge, UK.

Akira Miyake and Priti Shah, editors. 1999. *Models of Working Memory*. Cambridge University Press, New York.

Philip Resnik. 1992. Left–corner parsing and psychological plausibility. In *Proceedings of COLING*, pages 191–197.

Julie Van Dyke. 2002. *Parsing as Working Memory Retrieval: Interference, Decay, and Priming Effects in Long Distance Attachment*. Ph.D. thesis, University of Pittsburgh, PA.

Shravan Vasishth, Richard L. Lewis, Hans Uszkoreit, and Rama Kant Agnihotri. 2004. Distinguishing distance and decay. Submitted.

Shravan Vasishth. 2003. Quantifying processing difficulty in human sentence parsing: The role of decay, activation, and similarity-based interference. In *Proceedings of the EucoCogSci conference*, Osnabrueck, Germany.