Quantifying Processing Difficulty in Human Language Processing

Shravan Vasishth (Potsdam, Germany)

May 1, 2005

Introduction

One goal common to human sentence processing theories is to develop a cross-linguistically applicable account of human parsing processes. There is much empirical evidence consistent with such theories, based on experiments involving diverse languages such as English, Japanese, Korean, Dutch, and German, among others. However, processing facts for many other languages are yet to be determined, and it is still unknown how such cross-linguistic theories fare when their predictions are applied to these other languages. Hindi[1] is a particularly interesting language for this purpose; although much is known about Hindi syntax, semantics, pragmatics, etc., currently there exists almost no experimentally

[1]Hindi, also known as Urdu, or Hindi-Urdu, is an Indo-Aryan language spoken primarily in South Asia; it has about 424 million speakers in India (source: 1991 Census of India, www.censusindia.net), and about 10 million in Pakistan (source: www.sil.org). Although Urdu and Hindi use different scripts, and even though political pressures have resulted in an increasingly divergent lexicon, the syntax is essentially indistinguishable. Accordingly, the present research is likely to be equally relevant to Urdu as to Hindi.

grounded work on Hindi sentence processing.[2] Hindi also has certain interesting properties which make it a useful test case for evaluating current processing models.

In this paper, I present the results of a self-paced experiment that investigates the different predictions of two theories of human sentence processing: Hawkins' Early Immediate Constituents (1994, 1998), and Gibson's Dependency Locality Theory (2000). I show that neither theory can adequately account for the Hindi data, and propose an alternative explanation in terms of empirically well-motivated assumptions about activation decay in memory.

Self-center embeddings

Self-center-embedding constructions (hereafter, SCEs) are grammatical structures in which a constituent occurs medially within a larger instance of the same kind of syntactic category. Examples from English are sentences like (1) (Sampson, 2001, 15,19,18). In the example the embedded clauses are enclosed in square brackets.

(1) Don't you find [that sentences [that people you know produce] are easier to understand]?

Contrary to a commonly held belief, SCEs occur suprisingly frequently in language (Roeck et al., 1982), particularly in head-final languages. They have been

[2] An exception is the research of Vaid and Pandit (1991) on agent-role interpretation in normal and aphasic Hindi speakers.

the subject of much psycholinguistic research, in Dutch, see (Dickey & Vonk, 1997), (Kaan & Vasić, 2000); (Konieczny, 2000); Dutch and German (Bach, Brown, & Marslen-Wilson, 1986); and Japanese (Lewis & Nakayama, 2001), (Babyonyshev & Gibson, 1999), (Uehara & Bradley, 1996), among other languages.

The reason that SCEs have attracted attention in sentence processing research is that such constructions (especially in head-final languages) necessarily overload SHORT TERM or WORKING MEMORY (see (Miyake & Shah, 1999) and references cited there).

It is easy to see that working memory is taxed during the processing of SCEs, particularly in head-final languages. Consider the examples in (2) from Dutch, German, Japanese, and Hindi.

(2) a. (dat) Aad Jantje de lerares de knikkers liet helpen opruimen
that Aad Jantje the teacher the marbles let help collect
'(that) Aad let Jantje help the teacher collect the marbles.'

b. (dass) die Männer haben Hans die Pferde füttern lehren
that the men have Hans the horses feed teach
'(that) the men have taught Hans to feed the horses.'

c. Keiko-ga Tadashi-ga Kenji-o kiraida-to omotteiru
Keiko-nom Tadashi-nom Kenji-acc hates-comp thinks
'Keiko thinks that Tadashi hates Kenji.'

d. Siitaa-ne Hari-ko kitaab khariid-neko kahaa
Sita-erg Hari-dat book buy-inf said
'Sita told Hari to buy a/the book.'

During the course of real-time processing of such sentences, certain memory processes must necessarily occur. Pre-theoretically, the noun phrases (NPs) must somehow be temporarily encoded and stored in working memory until one or more verbs necessitate the NPs' retrieval and subsequent integration with the verb(s).

The above observation about storage and retrieval/integration processes during sentence processing, along with the fact that working memory is resource bounded (Miller, 1956), suggests that inherent constraints on working memory are likely to be the source of processing difficulty (cf. (Christiansen & MacDonald, 1996)). This characterization raises several questions: what exactly is stored and encoded, and how? What factors affect retrieval and integration? What has a greater impact on memory overload, retrieval or integration? The answers to these questions go beyond sentence processing; they are fundamental to understanding human attention and have wide-ranging applications in any area concerned with the consequences of cognitive overload in humans, such as aphasiology, attention-related disorders, the design of time-critical systems, language learning, etc.

Research in cognitive psychology has provided a rich body of results regarding working memory that suggest answers to some of the basic questions. One plausible factor is its general resource-bounded nature (see, e.g., (Joshi, 1990), (Rambow & Joshi, 1994), (Lewis, 1996), (Cowan, 2001)). Other plausible candidates are phonological similarity (Baddeley, 1966), (Baddeley, Thompson, & Buchanan, 1975) (Gathercole & Baddeley, 1993), decay of stored items (Brown, 1958), (Peterson & Peterson, 1959) and interference (Waugh & Norman, 1965), specifically proactive interference (Müller & Pilzecker, 1900) and retroactive interference (Keppel & Underwood, 1962) among stored items.

Hawkins and Gibson separately define other constraints that are also derived (although less directly) from the psychology literature and are ultimately related to working memory. In Hawkins' EIC theory, the main factor is the number of words the perceiver needs to see/hear in order to recognize a phrasal category: the more the number of words per phrase, the harder a sentence is to process. This translates to a complexity metric that yields specific predictions about particular sentence structures across languages; the complexity metric can be regarded as quantifying the strain on a resource-bounded working memory. A simple example is the Hindi sentence *hari-ne kitaab khariidii*, 'Hari bought a book' (literally: Hari-erg book bought). Here, the constituent VP 'book bought' contains only two words at the moment that it is completed (at the verb). Inserting an adverb like 'yesterday' between the object and the verb would, however, increase the number of words in the constituent to three. According to Hawkins' metric, this results increased cost of building the constituent, presumably when the verb is processed.

By contrast, Gibson's model quantifies the constraints on a limited working memory in terms of (*inter alia*) the number of new discourse referents introduced so far. Under this view, integrating an argument with a verb is partly a function of the number of discourse referents that intervene. Thus, increased processing difficulty is predicted at the verb *went* in a sentence like *The man who was wearing a hat went home*, compared to the case where the relative clause is absent.

The distance hypothesis: An evaluation

I will refer to as the DISTANCE HYPOTHESIS the above claim in EIC and DLT that increasing distance between certain arguments or between dependents and heads results in increased processing difficulty.

As discussed above, EIC predicts an increase in processing difficulty at the innermost verb if an adverb intervenes between the final NP and verb. DLT, on the other hand, assumes that inserting an adverb should have no effect on processing complexity (since no discourse referent is introduced), but Gibson also suggests that it could result in increased processing difficulty (if other factors, other than discourse status, affect distance (Gibson, 2000, 107)).

These predictions suggest an obvious experiment, which I describe next.

Subjects, Method, and Materials

Sixty undergraduate and graduate students at Jawaharlal Nehru University, New Delhi participated in the experiment. Each subject was paid one hundred Rupees (equivalent approximately to two US dollars in September 2001) for participating.

Four lists were prepared in a counterbalanced, Latin Square design, and 44 fillers were inserted between 32 target sentences in pseudorandomized order. The sentence types are as shown in (3) (a full list of the stimulus sentences is presented in (Vasishth, 2003b)).

(3) a. A. **Single embedding, no intervening adverb**:

Siitaa-ne Hari-ko [kitaab-ko khariid-neko] kahaa
Sita-erg Hari-dat book-acc buy-inf told
'Sita told Hari to buy the book.'

b. B. **Single embedding, adverb intervenes**:

Siitaa-ne Hari-ko [kitaab-ko **jitne-jaldi-ho-sake** khariid-neko]
Sita-erg Hari-dat book-acc as-soon-as-possible buy-inf
kahaa
told
'Sita told Hari to buy the book as soon as possible.'

c. C. **Double embedding, no intervening adverb**:

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

d. D. **Double embedding, adverb intervenes**:

Siitaa-ne Hari-ko Ravi-ko [kitaab-ko **jitne-jaldi-ho-sake**
Sita-erg Hari-dat Ravi-dat 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.'

In each of the test sentences, all but the final NPs were proper names; the final NP was always an inanimate common noun, such as 'book' or 'letter.' The fillers consisted of various syntactic structures, such as relative clauses, medial gapping constructions, simple declaratives, and sentences with that-clauses. This was a self-paced moving window reading task (Just, Carpenter, & Woolley, 1982). The white spaces between each phrase/word in (3) correspond to the segmentation in the self-paced reading task.

Results

A subject (F1) and item (F2) contrast analysis for single embeddings (Conditions A versus B) is shown in Tables 1 and 2, and are summarized graphically in Figure 1. The innermost verb showed no significant difference in RT for the two conditions. Since Figure 1 suggests that the third NP position may have significantly different RTs in the two conditions, a contrast analysis for this position was done as well. The results show a significantly shorter RT in the sentence with the adverb present, but this was only true for the by-subjects analysis.

The analyses for double center embeddings (Conditions C versus D) are summarized in Tables 3 and 4, and are shown graphically in Figure 2. The contrast analysis for the innermost verb showed a significantly[3] shorter RT in the sentence with the adverb present, but only in the by-subjects analysis. Since Figure 2 suggests that the fourth NP position may have significantly different RTs in the two conditions, a contrast analysis for this position was done as well. However, no significant difference was found at the fourth NP.

[3] Although a p-value of 0.02 is not particularly compelling evidence, this result has been replicated in other experiments (Vasishth, 2003a; Vasishth & Lewis, 2003).

Reading time	F1(1,51)	p-value	F2(1,31)	p-value
Raw	5.91	0.0189	3.48	0.0768

Table 1: Conditions A vs. B; contrast analysis for NP3

Reading time	F1(1,51)	p-value	F2(1,31)	p-value
Raw	2.77	0.1026	2.99	0.0989

Table 2: Conditions A vs. B; contrast analysis for V2

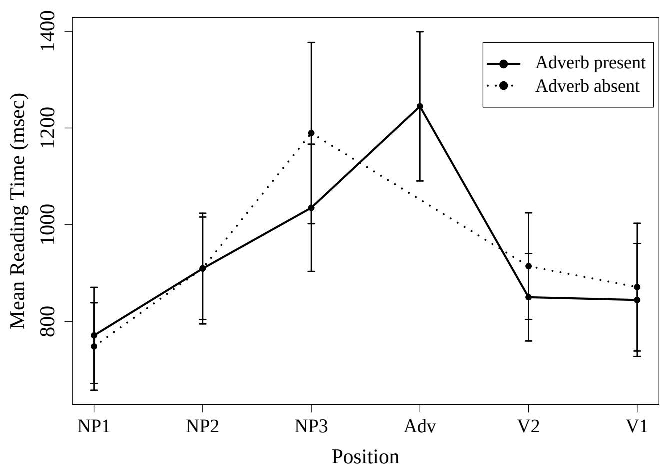

Figure 1: Conditions A vs. B; raw RTs, with 95% confidence intervals

Reading time	F1(1,51)	p-value	F2(1,31)	p-value
Raw	2.32	0.1339	1.71	0.2053

Table 3: Conditions C vs. D; contrast analysis for NP4

Reading time	F1(1,51)	p-value	F2(1,31)	p-value
Raw	5.08	0.0288	2.21	0.1526

Table 4: Conditions C vs. D; contrast analysis for V3

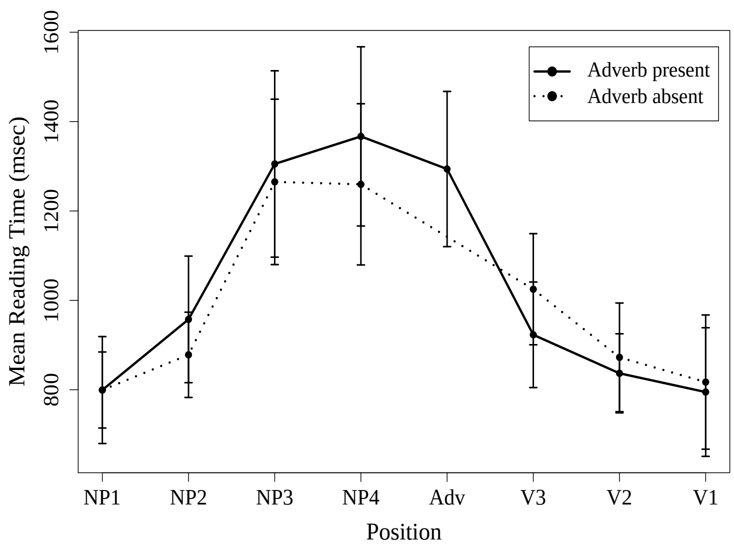

Figure 2: Conditions C vs. D; raw RTs, with 95% confidence intervals

Discussion

Single embeddings failed to show a significant difference at the innermost verb when the adverb was present. This is an inconclusive null result. But if this were conclusive it would be consistent with one instantiation of the DLT (Gibson, 2000, 107), which predicts that adverb-insertion should have no effect. However, in double embeddings, there was a significantly *shorter* RT at the innermost verb, and this is inconsistent with the EIC and DLT distance hypothesis.

One explanation for the speedup (another explanation, in terms of gradual acceleration in reading speed, is discussed and rejected in (Vasishth, 2003b)) is that adverbs may generally be read faster than NPs, and therefore any spillover from an adverb to a subsequent verb would be less than a spillover from an NP to a subsequent verb. However, the data suggest that this is incorrect. Consider the reading times for the final NP and the adverb in the single embedding's Condition B. A comparison of the reading times at these two positions shows that there is a statistically significant *slowdown* at the adverb following the final NP ($F1(1,51)= 9.4357$, $p = 0.0035$; $F2(1,23)= 8.4188$, $p = 0.008824$). Moreover, in the double embeddings, although the reading time for the adverb appears to be faster than that of the final NP, the effect was not significant ($F1(1,51) = 0.5054$, $p = 0.48057$; $F2(1,23)= 0.9303$, $p = 0.346295$). Finally, independent evidence from self-paced reading and eyetracking studies suggests that adjuncts are in general read slower than arguments, not faster (Clifton, Speer, & Abney, 1991).[4]

[4] See (Vasishth, 2003a) for a more recent self-paced reading study which found that intervening items other than adverbs also result in a speedup at the verb. At the very least, this suggests

Thus, the spillover argument cannot account for the double embedding facts. One might argue that adverbs are longer than NPs (at least in this experiment), and so factoring out word length might change this result. However, as discussed in (Vasishth, 2003b), word length does not appear to have much effect on reading time in these particular experiments. Even if, counterfactually, word length did affect reading time, the spillover argument is that a speedup at the adverb is passed on to the verb. But since there is no speedup at the adverb (rather, there is a slowdown in the single embedding case), the word length issue is orthogonal to the present question.

A computational model of sentence processing

Although both DLT and EIC correctly characterize important generalizations about human parsing processes, the speedup effect presented here does not support these models. I discuss next an alternative approach that does explain existing results as well as the speedup effect. The precise details of the model are available in (Vasishth & Lewis, 2004), and (Lewis & Vasishth, 2005).

The key idea is that instead of computing distance in terms of discourse referents or the number of words per constituent, processing difficulty is quantified in terms of activation decay. The question then reduces to finding a principled basis

---

that the speedup is not due to an adverb being present but due to some other factor. Cf. the eyetracking study reported in (Vasishth, Cramer, Scheepers, & Wagner, 2003) which found a speedup in a German center embedding eyetracking study only when adverbs intervened. I am currently investigating the possibility that this divergence between the self-paced reading and eyetracking experiments is due to the different methodologies used (Sturt, Scheepers, & Pickering, 2002).

for defining rate of decay.

An empirically well-motivated candidate is the activation decay mechanism incorporated in the cognitive architecture ACT-R (Anderson, Bothell, Byrne, & Lebiere, 2002). In ACT-R, an item i in permanent memory is assumed to have a base-level activation B_i; reflects the log-odds that an item will be needed (Anderson & Schooler, 1991). In ACT-R the odds that a fact will be needed decays as a power function of how long it has been since it has been used.

In addition, the effect of multiple retrievals is quantified according to the equation shown below. Here, t_j is the time elapsed since the j^{th} retrieval of the item i. The parameter d determines the effect of each retrieval on B_i; a default value, $d = 0.5$, has emerged as a consequence of modeling in a large range of applications (Anderson et al., 2002).

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

Retrieval time T_i of item i is a function of the activation level, and total reading time depends partly on retrieval time.

(5) $$T_i = Fe^{-A_i}$$

In addition to the above constraints, if, following existing research (Resnik, 1992), we assume that the human sentence parsing mechanism is essentially a left-corner parser, we can straightforwardly account for the speedup effect, in addition to other well-known results in the literature (see (Lewis & Vasishth, 2005; Vasishth, Drenhaus, Saddy, & Lewis, 2005)).

In essence, the speedup follows as a consequence of the fact that as soon as the final NP in the double center embedding is processed, an embedded verb phrase (VP) is predicted. This VP, which is constructed on the fly, has a base-level activation that is decaying as time passes. In the case where no adverb intervenes between the verb and the preceding NP, the VP is integrated into the structure built incrementally so far. Assume that this integration, which is partly a function of the retrieval time required for the VP, takes time t.

In the case where an adverb intervenes, at the NP the embedded VP is constructed as sketched above, and then the adverb is processed. Part of the adverb processing involves retrieving this newly created VP and being integrated with the VP. When the verb is processed next the VP must be retrieved a second time. This second retrieval occurs faster, in time $t' < t$, than in the case of the no-intervening adverb condition. The decay that the VP undergoes as the adverb is processed is offset by its increase in activation due to the extra retrieval at the verb.

In sum, apart from furnishing better empirical coverage, the activation decay mechanism allows us to move away from a characterization of processing difficulty as a *linguistically* defined property of the parsing mechanism, and towards a characterization directly in terms of working memory constraints. The present approach therefore entails that human sentence processing is better construed as a general information processing task that the human brain engages in, rather than as an exclusively linguistic activity subject to linguistically defined constraints like constituent to word ratios or the number of intervening discourse referents. The critical difference between the DLT and EIC class of approaches and the present one is that the latter generalizes to different kinds of information processing tasks, but the former does not. Apart from the fact that it does a better job of capturing the known behavioral facts, the cognitive-constraint approach is therefore *a priori* independently motivated

Acknowledgements

This research is part of my doctoral dissertation work at The Ohio State University's Linguistics Department (Vasishth, 2003b). It was funded by the Department of Linguistics, the Department of Computer and Information Science, and the Center for Cognitive Science, The Ohio State University. I am very grateful to Richard L. Lewis and Shari Speer, my advisors at Ohio State, and Ayesha Kidwai (Jawaharalal Nehru University), who graciously provided logistical help for running the experiments.

References

Anderson, J., Bothell, D., Byrne, M., & Lebiere, C. (2002). *An integrated theory of the mind.* (submitted to Psychological Review)

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2(6)*, 396–408.

Babyonyshev, M., & Gibson, E. (1999). The complexity of nested structures in Japanese. *Language*, *75(3)*, 423–450.

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes, 1(4)*, 249–262.

Baddeley, A. (1966). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*.

Baddeley, A., Thompson, N., & Buchanan, M. (1975). Word-length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*, 173–189.

Christiansen, M. H., & MacDonald, M. C. (1996). *Processing of recursive sentence structure: Testing predictions from a connectionist model.* (MS)

Clifton, C., Speer, S., & Abney, S. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, *30*, 251–271.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24(1)*, 87–114.

Dickey, M. W., & Vonk, W. (1997). *Center-embedded structures in Dutch: An on-line study.* (MS)

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Hove, UK: Lawrence Erlbaum Associates.

Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium.* Cambridge, MA: MIT Press.

Hawkins, J. A. (1994). *A Performance theory of order and constituency.* New York: Cambridge University Press.

Hawkins, J. A. (1998). Some issues in a performance theory of word order. In A. Siewierska (Ed.), *Constituent order in the languages of Europe.* Berlin: Mouton de Gruyter.

Joshi, A. K. (1990). Processing crossed and nested dependencies: An automaton perspective on the psycholinguistic results. *Language and Cognitive Processes*, *5(1)*, 1–27.

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

Kaan, E., & Vasić, N. (2000). *Cross-serial dependencies in Dutch: Evidence for a discourse-based metric?* (Poster presented at CUNY Sentence Processing Conference, University of California, San Diego)

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1*, 153–161.

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

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, *25(1)*, 93–115.

Lewis, R. L., & Nakayama, M. (2001). Syntactic and positional similarity effects in the processing of Japanese embeddings. In M. Nakayama (Ed.), *Sentence Processing in East Asian Languages* (pp. 85–113). Stanford, CA.

Lewis, R. L., & Vasishth, S. (2005). *An activation-based model of sentence processing as skilled memory retrieval.* (Accepted for publication in Cognitive Science)

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.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory.* New York: Cambridge University Press.

Müller, G. E., & Pilzecker, A. (1900). *Experimentelle Beiträge zur Lehre vom*

Gedächtnis (Experimental contributions on the theory of memory). Leipzig: J. A. Barth.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, *61*, 12-21.

Rambow, O., & Joshi, A. K. (1994). A processing model for free word order languages. In J. C. Clifton, C., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing.* Hillsdale, NJ: L. Erlbaum.

Resnik, P. (1992). Left–corner parsing and psychological plausibility. In *Proceedings of COLING* (pp. 191–197).

Roeck, A. D., Johnson, R., King, M., Rosner, M., Sampson, G., & Varile, N. (1982). A myth about center-embedding. *Lingua*, *58*, 327–340.

Sampson, G. (2001). *Empirical linguistics.* London: Continuum.

Sturt, P., Scheepers, C., & Pickering, M. (2002). Syntactic ambiguity resolution after initial misanalysis: The role of recency. *Journal of Memory and Language*, *46*, 371–390.

Uehara, K., & Bradley, D. (1996). The effect of *-ga* sequences on processing Japanese multiply center-embedded sentences. In *11th Pacific-Asia Conference on Language, Information, and Computation.*

Vaid, J., & Pandit, R. (1991). Sentence interpretation in normal and aphasic Hindi speakers. *Brain and Language*, *41*, 250–274.

Vasishth, S. (2003a). Quantifying processing difficulty in human sentence parsing: The role of decay, activation, and similarity-based interference. In *Proceedings of the European Cognitive Science Conference (Osnabrück).* Lawrence Erlbaum.

Vasishth, S. (2003b). *Working memory in sentence comprehension: Processing Hindi center embeddings.* New York: Garland Press. (Published in the series Outstanding Dissertations in Linguistics, edited by Laurence Horn)

Vasishth, S., Cramer, I., Scheepers, C., & Wagner, J. (2003). Does increasing distance facilitate processing? In *Proceedings of the 16th Annual CUNY Sentence Processing Conference.* MIT, Cambridge, MA.

Vasishth, S., Drenhaus, H., Saddy, D., & Lewis, R. (2005). Processing negative polarity. In *Proceedings of the CUNY Sentence Processing Conference.* University of Arizona.

Vasishth, S., & Lewis, R. (2003). The role of decay and activation in human sentence processing. In *Proceedings of the AMLaP Conference.* Glasgow, Scotland.

Vasishth, S., & Lewis, R. (2004). Modeling sentence processing in ACT-R. In *Proceedings of the Workshop on Incremental Parsing, 42nd Annual Meeting of the Association for Computational Linguistics.* Barcelona, Spain.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, *72*, 89–104.