Length vs. Structural Complexity in Sentence Comprehension in Aphasia

Susan Curtiss and Catherine A. Jackson
University of California-Los Angeles, Los Angeles, California

Daniel Kempler and Wayne R. Hanson
Veterans Administration Medical Center, Sepulveda, California

E. Jeffery Metter
Veterans Administration Medical Center, Sepulveda, California and
University of California-Los Angeles, Los Angeles, California

The Token Test (hereafter TT) has been shown sensitive and reliable for measuring auditory language comprehension (e.g., DeRenzi and Vignolo, 1962; Kertesz, 1982; McNeil and Prescott, 1978; Orgass and Poeck, 1966). The test has been found to detect even "latent" aphasia in apparently recovered individuals (Boller and Vignolo, 1966). A number of different versions of the TT (e.g., DeRenzi and Vignolo, 1962; Spreen and Benton, 1969; Spellacy and Spreen, 1969) have been developed and used to detect and diagnose the presence and severity of aphasia (Gaddes and Crockett, 1973; Swisher and Sarno, 1968; Poeck, Kerschensteiner, and Hartje, 1972; DeRenzi and Vignolo, 1962). What exactly the test is testing, however, is open to some question.

The test is typically divided into five sections, four of them consisting of sentences of increasing length (4, 5, 8, and 10 words in Parts I - IV, respectively). All of these sections involve the same basic syntactic and semantic structure -- a simple imperative sentence with an imperative verb followed by a Noun Phrase (or NP and NP combination) in which a noun is modified by one or more attributive adjectives, as illustrated in (1) - (4) below:

(1) Touch the red circle. (from Part I)
(2) Touch the large red circle. (from Part II)
(3) Touch the red circle and the yellow rectangle. (from Part III)
(4) Touch the small blue circle and the small red circle (from Part IV)

Only the last section of the test contains items that differ in semantic and syntactic structure, such as (5) and (6):

(5) Touch the blue circle with the red rectangle.
(6) Pick up all the circles except the yellow one.

This section contains sentences which vary in length from 8-12 words and includes a variety of semantic structures (e.g., quantifiers, locatives, adverbs, and logical connectives) and syntactic structures (e.g., conjoined clauses, extraposed prepositional phrases, and adjunct clauses). Three of the parts use an array of 10 tokens; the other two use a 20-item array.

Accounts of the performance patterns of various subject groups on the TT suggest that it may define different deficiencies in different populations. Tallal, testing language-impaired children (1975), and Zaidel, testing split-brain and hemispherectomy patients (1977), for example, found that their subjects performed worse on part IV than on part V and argue that auditory verbal short term memory was the more crucial factor affecting performance in their subjects. In other studies (e.g., Whitaker and Noll, 1972, and Swisher and Sarno, 1969) both normals and aphasic and nonaphasic brain-damaged subjects performed worst on part V, suggesting that for many individuals, children and adults alike, linguistic complexity may affect TT performance as much as, or more than, sentence length. What is more, several studies report large interitem performance differences on part V (Whitaker and Noll, 1972; Poeck et al., 1974; Mack and Boller, 1979), indicating that part V does not test a homogeneous set of linguistic abilities.

Everything considered, the TT tests the integrity of (receptive) performance and processing factors such as the ability to handle largely nonredundant information, array size, and increasing memory load as much as it tests comprehension of linguistic structure per se. However, the TT does not systematically contrast length with linguistic complexity. For example, there are no short, linguistically complex items on the test. Therefore, while it has been reported that aphasic subjects show impaired performance relative to normal and to right-brain-damaged adults, it has not been established whether factors such as sentence length or structural complexity negatively affect TT performance in aphasic subjects. We constructed a modified version of the TT to investigate the relative effects of sentence length and structural complexity on sentence comprehension in aphasia. We presented to aphasic patients actual TT sentences along with syntactically complex sentences of the same length that we devised. In this way, we systematically compared the effect of sentence length with linguistic complexity. Given that aphasia involves deficits in linguistic performance, our predictions were (1) that structural complexity would impair comprehension even at the shortest sentence lengths, and (2) that there would be a simple inverse relationship between sentence length and correct performance, such that the greater the length, the poorer the performance.

EXPERIMENT 1

Method

Subjects. Subjects were 26 aphasic adults (21 men and 5 women). All were right-handed, native speakers of English, ranging in age from 37 to 72 years with a mean age of 60. The subjects were seen as part of a larger study involving both chronic and acute patients; they ranged from 1 to 81 months post onset of aphasia (mean = 28.12 months). The subjects were classified into types on the basis of their performance on the Western Aphasia Battery (WAB; Kertesz, 1982): 11 anomic, 7 Broca's, 4 Conduction, 3 Wernicke's and 1 Transcortical Sensory. Aphasia Quotients (AQ's) ranged from 19 to 98 (mean = 69.31). The distribution of AQ's was: 80 to 100 -- 12 subjects; 60 to 80 -- 5 subjects; 40 to 60 -- 5 subjects; below 40 -- 4 subjects. Ten normal adults (5 men and 5 women) also participated in the study. All were native English speakers, with a mean age of 60.1 (range = 51 to 74) years.

Procedures. Using the prescribed format of the TT, we administered a modified version consisting of four parts. Each part had items which were either "simple" or "complex" (i.e. embodied structural complexity). Part 1 consisted of six 4-word sentences -- three simple items which were actual TT items, and three complex items which we devised. Part 2 consisted of six 5-word sentences -- again, three simple TT items and three complex items which we constructed. Part 3 sentences were 8 words in length, and part 4 sentences were 10 words in length. In these two sections, both the simple and complex items were actual TT items (with one of the complex items slightly modified to correspond exactly in length to the other items). The entire set of items is presented in Appendix 1. As with other versions of the TT, those parts with items mentioning size adjectives were administered with an array of 20, and the other used an array of 10.

In devising our version of the TT, we constructed sentences that were complex from the standpoint of their formal structural representations as defined within the theory of grammar known as Government Binding (GB) theory. In GB theory there are several levels of representation: D-structure (underlying structure), S-structure (the level which is the output of the transformational component) and Logical Form (LF) (the level which maps syntactic form onto a semantic interpretation). (See Chomsky, 1981 and Van Riejimsdijk and Williams, 1986 for details.) The sentences we devised were more complex than same-length sentences from the TT in that our items embodied one or both of the following parameters, neither of which were part of any of the TT sentences:

(1) WH-movement or quantifier raising (QR). Some constituents can be generated in one sentence position (node) at D-structure and moved to another position (node) either at S-structure of LF within specified constraints of the theory. Such movement (in contrast with movement of clauses, for example) leaves behind the category dominating the moved element (e.g., the noun phrase (NP) in 8a') in its structural representation, and this now empty category is marked with a "trace" which is coindexed with the moved element. Movement in the syntax (in the transformational component) accounts for the appearance of WH constituents at the front of their clauses in relatives and at the front of the highest S in WH-questions. WH-movement at LF accounts for the grammaticality and interpretation facts for multiple WH-questions, such as 'Which boy read which book?' QR occurs at LF and accounts for the grammaticality and interpretation facts regarding pronoun-quantifier antecedent relations and quantifier scope.

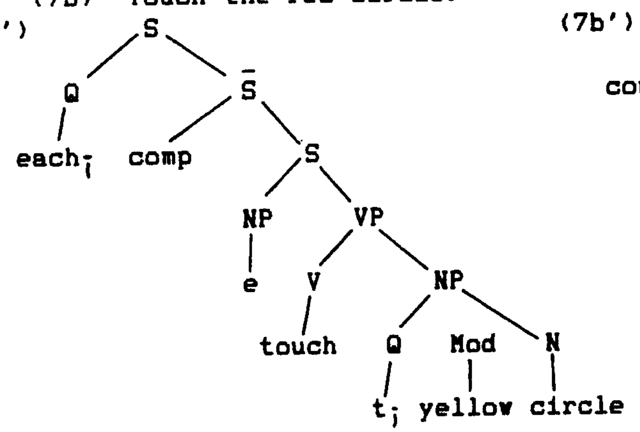
(2) A sentence (S) embedded within another S. (here, a relative clause). In relative clauses, the relativized (WH) constituent is moved to the COMP (complementizer) position at the front of its clause, leaving a trace of movement behind, after which it may be deleted under conditions of recoverability. A relative clause, then, is an embedded S which also involves WH-movement, thus embodying both parameters (1) and (2).

By our definition of complexity, then, sentence (7a) below, which has an LF structure like (7a') is complex as opposed to (7b) which has an LF representation like (7b') and is simple.

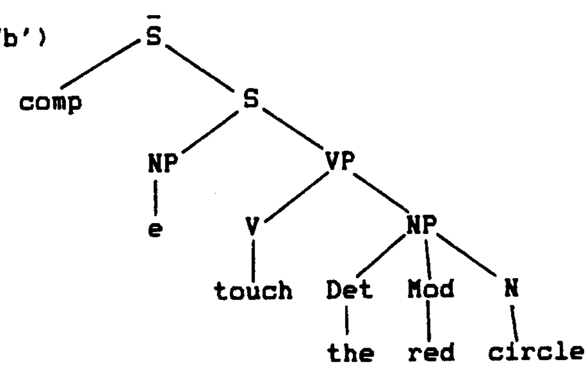
(7a) Touch each yellow one.

(7b) Touch the red circle.

(7a')

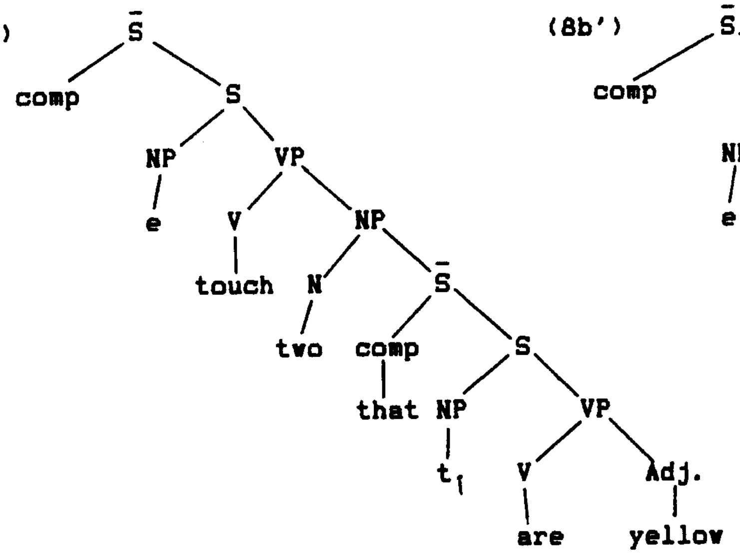


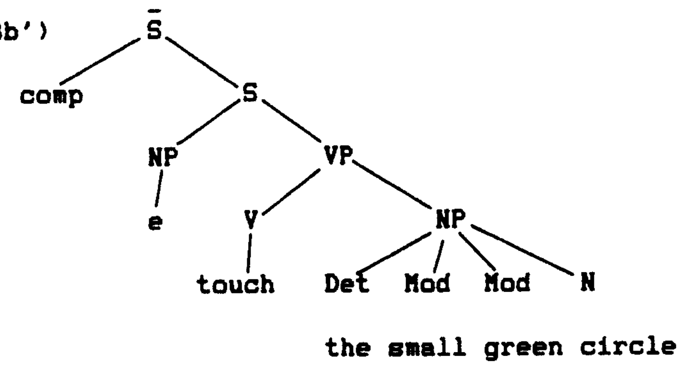
(7b')



Similarily, sentence (8a) which has an S-structure representation as in (8a') is complex, but (8b) which has an S-structure representation as in (8b') is simple.

(8a) Touch two that are yellow
(8b) Touch the small green circle.

(8a')


(8b')


Results

The performance of the 26 aphasic subjects is shown in Tables 1 and 2. Table 1 shows the total number of correct responses of the total number possible; Table 2 shows the mean number correct. One control subject made a single error.

Table 1. Overall number of sentences comprehended (78 possible).

	Length in words			
	4	5	8	10
Type of S:				
Simple	71 (91%)	64 (82%)	47 (60%)	36 (46%)
Complex	60 (76%)	50 (64%)	39 (50%)	38 (49%)

For the aphasic patients, a two-way analysis of variance indicated a significant effect for length--$F(6, 150) = 19.41$, with $p < .01$. In contrast, using TT items from published versions of the test as the complex longer items, no consistent effect of complexity was found--$F(1, 50) = 1.49$, $p > .05$). There was, however, a significant difference between the simple and complex items at the two shortest lengths ($p < .01$, two-tailed <u>t</u>-test. There was also a significant length/complexity interaction (for $F(3, 140) = 2.78$, $p < .05$). Prediction 2 (i.e., the longer the sentence, the poorer the performance will be) was supported by these results, but prediction 1 (that structural linguistic complexity would consistently impair performance) was supported only in part (i.e., for the two shortest lengths).

Table 2. Mean number of correct responses by sentence type and length.

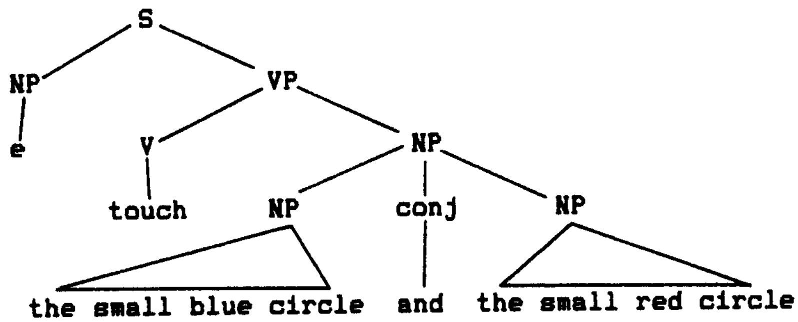
	Length in words			
	4	5	8	10
Type of S:				
Simple	2.73	2.46	1.81	1.38
Complex	2.31	1.92	1.50	1.46

We were surprised by our results indicating that the long-complex items were not significantly harder to comprehend than the long-simple items, especially the finding that performance on the 10-word long-complex items was as good as on the 10-word long-simple items. There were three plausible explanations. First, it could have been the case that the particular semantic and syntactic structure of the long-simple items allowed for no reduction in what must be held in memory, making them as or more difficult to process than the long-complex items. The structural variation within the complex items limits the concatenation of elements to be held in short term memory. However, in the simple items at the longest length there are two attributive adjectives for each noun. Thus, along with the shape of the token named by the noun, there are three distinct items (which are not related, reducible or chunkable) to be stored in memory for each noun phrase. In contrast, the 10-word complex items have at most a single attributive adjective modifying each noun and have the additional feature that at least some of the lexical items combine to form a semantic or logical relationship, such as the temporal relationship between the actions in the case of the "before" item, or the conditional relationship between the actions in the "if" item. In these two examples, there are also two "action-object" relationships or units per sentence (e.g., "touch [noun phrase]"). The logical or semantic structure between parts of the sentence in the case of the complex items may thus allow groups of words to be processed as single units, making the information more reducible, thereby aiding its short-term memory processing.

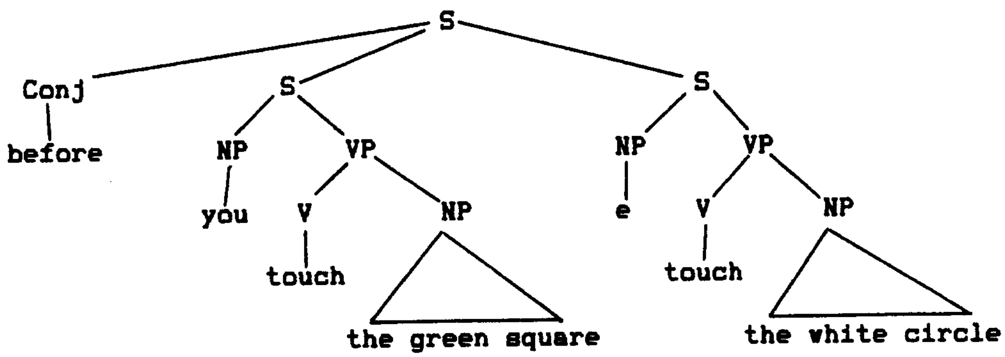
A second possible explanation was that the long complex items taken from the TT were actually not more complex than the long simple items. The short complex sentences (the items we devised) were designed to embody particular syntactic parameters which would make them structurally more complex than their simple counterparts. The long-complex items, in contrast, were essentially only structurally different from, not more complex than, the same-length simple items, as shown in example (9a) (simple) and (9b) complex) below, which have the S-structure representations of (9a') and (9b'), respectively:

(9a) Touch the small blue circle and the small red circle.
(9b) Before you touch the green square touch the white circle.

(9a')



(9b')



The 10 (and 8)-word complex sentences of this experiment, then, merely involved conjoined sentences rather than conjoined NP's, or a noun phrase and a prepositional phrase, rather than two noun phrases.

A third possibility is that the ability to attend to the more complex stimulus array used with the 10-word simple items (as compared to the simpler array used with the 10-word complex items) may have impaired performance, confounding the effects of length and/or grammatical complexity.

We therefore conducted an additional experiment to decide between the first two possibilities and to eliminate the confounding effects of stimulus array size. Our prediction was that when complexity was defined along specific grammatical parameters, it could be shown to impair comprehension.

EXPERIMENT 2

Method

Subjects. Five of the aphasic subjects who participated in Experiment 1 and six additional patients selected in the same manner as for Experiment 1 served as subjects for this experiment. These patients ranged from 1 to 70 months post onset (mean = 21 months), and were classified as follows: 4 anomic, 3 Broca's, 3 conduction, and 1 Wernicke's type. They also demonstrated a range in severity as measured by their WAB AQ, ranging from 29 to 95 (mean = 70.36). The distribution of their AQ's was as follows: 80 to 100 -- 4 patients; 60 to 80 -- 4 patients; 40 to 60 -- 1 patient; and below 40 -- 2 patients. The same 10 normals were used as controls as in Experiment 1.

Procedure. New 10-word complex items were constructed so that all had syntactic structures which involved an embedded sentence and WH and/or quantifier movement. Furthermore, they all involved movement from object position, thereby leaving object "traces," structures which have been reported to be particularly problematic for at least some aphasic individuals (Grodzinsky, 1986).

4. Touch the circle that the small green circle is beside.
5. Touch each square the small blue square is next to.
6. Touch both circles that the large red circle is between.

These structures were chosen because they embodied structural complexity as defined above (see section 3.0), and thus were parallel to the complex items devised for the shorter items. In addition to the changes in the stimulus items, the array presented for both 10-word simple and complex items was the 20-token array (now appropriate for all 10-word items, because the new complex items, like the simple items, contained two adjectives modifying a noun). Patients new to the study were given the 4 - 8 word items of Experiment 1, the original 10-word simple items, and the new 10-word complex items; subjects from Experiment 1 were given only the new 10-word items.

Results

The control subjects made no errors. One of the new aphasic subjects also made no errors on the test. Results for the 11 aphasic subjects indicated that there was no significant difference between performance on simple and complex items for the aphasic subjects as a whole at the 10-word length.

However, on closer inspection, we found that the performance of most of the patients could be divided into two groups. The first group, the syntax sensitive group (4 subjects), were those patients whose performance was consistently worse on the complex items, both for the 10-word length and for the battery as a whole. The length sensitive group (4 subjects) were those patients who did poorly on all 10-word items, and whose performance level decreased with increasing sentence length on both simple and complex items. Two additional patients showed no clear effect of either length or complexity. The totals of the scores of patients in the two subgroups for simple versus complex items (both for the 10-word length and all lengths) are given in Table 3.

Table 3. Subgroup performance for Experiment 2.

	10-word items		All items	
	Simple (out of 12)	Complex (out of 12)	Simple (out of 48)	Complex (out of 48)
Syntax sensitive group (N = 4)	10 (83%)	5 (42%)	44 (92%)	33 (69%)
Length sensitive group (N = 4)	1 (10%)	1 (10%)	26 (54%)	21 (44%)

Discussion

In Experiment 1 and 2, we found that both syntactic complexity and sentence length affected sentence comprehension in our group of aphasic subjects. In general, our patients showed an effect for complexity at the shortest lengths. What differentiated patients from each other in both experiments was their sensitivity to sentence length; i.e., their ability to process relatively long sentences. It appears that certain patients have a significant linguistic short term memory impairment, such that sentence length alone can sufficiently affect sentence processing so as to minimize the importance of other parameters, including linguistic structure. Results from other sentence comprehension studies have also noted such patients (e.g., Goodglass, Blumstein, Hyde, Green, Gleason, and Statlender, 1979). However, other aphasic patients show a more strictly structural linguistic deficit; and when syntactic complexity is defined and tested in a systematic way (as in Experiment 2), a clear effect for structural complexity in comprehension is revealed for this second group. Since difficulty in sentence comprehension may result from distinct underlying impairments, these findings show the importance of evaluating the comprehension abilities of aphasic individuals on a case-by-case basis to determine exactly which parameters affect their performance.

Our findings also suggest that type of aphasia (determined on the basis of clinical criteria) may be unrevealing regarding the character of the particular comprehension impairment(s) involved. In Experiment 1, patients classified as anomic or conduction demonstrated better performance overall; this is consistent with the clinical characterization of these types of aphasia as being somewhat milder than other types in terms of overall impairment. However, in neither experiment do subgroups of patients defined on the basis of aphasia type perform homogeneously. In Experiment 2, for example, neither the length nor syntax subgroup is homogeneously comprised of a single aphasia type. The length sensitive group was composed of two anomic individuals, and one patient each with conduction and Broca's aphasia, while the syntax sensitive group had one individual each with anomic, Wernicke's, conduction, and Broca's aphasia. Thus, our findings support the position that the presence of a linguistic (syntactic) impairment in aphasia does not relate in any clear way to the classical syndromes described in the clinical literature (e.g., Mack, 1981; Goodglass and Menn, 1985; Heeschen, 1985), and that both grammatical and nongrammatical factors can contribute to comprehension impairment across aphasia classification boundaries.

REFERENCES

Boller, F. and Vignolo, L.A. Latent sensory aphasia in hemisphere-damaged patients: An experimental study with the Token Test. Brain, 89, 815-830, 1966.

Chomsky, N. Lectures on Government and Binding. Dordecht: Foris, 1981.

DeRenzi, E. and Vignolo, L.A. The Token Test: A sensitive test to detect receptive disturbances in aphasics. Brain, 85, 665-678, 1962.

Gaddes, W. and Crockett, D. The Spreen-Benton aphasia tests: Normative data as a measure of normal language development. Research Monograph No. 25, Department of Psychology, University of Victoria, Victoria, B.C., 1973.

Goodglass, H., Blumstein, S.E., Gleason, J.B., Hyde, M.R., Green, E. and Statlender, S. The effects of sentence encoding on comprehension in aphasia. Brain and Language, 7, 201-209, 1979.
Goodglass, H. and Menn, L. Is agrammatism a unitary phenomenon? In M.L. Kean (Ed.), Agrammatism. New York: Academic Press, 1985.
Grodzinsky, Y. Language deficits and the theory of syntax. Brain and Language, 27, 135-159, 1986.
Heeschen, C. Agrammatism versus paragrammatism: A fictitious opposition. In M.L. Kean (Ed.), Agrammatism. New York: Academic Press, 1985.
Huang, C. Logical relations in Chinese and the theory of grammar. MIT Ph.D. dissertation, 1982.
Mack, J. The comprehension of locative prepositions in nonfluent and fluent aphasia. Brain and Language, 14, 81-92, 1981.
Mack, J. and Boller, F. Components of auditory comprehension: Analysis of errors in a revised TT. In F. Boller and M. Dennis (Eds.), Auditory Comprehension: Clinical and Experimental Studies with the TT. New York: Academic Press, 1979.
May, R. Logical Form: Its Structure and Derivation. Cambridge, Mass: The MIT Press, 1981.
McNeil, M.R. and Prescott, T.E. Revised Token Test. Baltimore: University Park Press, 1978.
Orgass, B. and Poeck, K. Clinical validation of a new test for aphasia: An experimental study on the Token Test. Cortex, 2, 222-243, 1966.
Peock, K., Kerschensteiner, M. and Hartje, W. A quantitative study on language understanding fluent and nonfluent aphasia. Cortex, 299-304, 1972.
Spellacy, F.J. and Spreen, O. A short form of the Token Test. Cortex, 5, 390-397, 1969.
Spreen, O. and Benton, A.L. Neurosensory Center Comprehensive Examination for Aphasia, Edition A. Manual for Instructions. Neuropsychology Laboratory, University of Victoria, 1969.
Swisher, L.P. and Sarno, M.T. Token Test scores of three matched patient groups: left brain-damaged with aphasia; right brain-damaged without aphasia; non-brain-damaged. Cortex, 5, 264-273, 1969.
Tallal, P. Perceptual and linguistic factors in the language impairment of developmental dysphasics: An experimental investigation with the Token Test. Cortex, 11, 196-205, 1975.
van Riemsdijk, H. and Williams, E. Introduction to the Theory of Grammar. Cambridge, Mass: The MIT Press, 1986.
Whitaker, H.A. and Noll, J.D. Some linguistic parameters of the Token Test. Neuropsychologia, 10, 395-404, 1972.
Zaidel, E. Unilateral auditory language comprehension on the Token Test following cerebral commissurotomy and hemispherectomy. Neuropsychologia, 15, 1-17, 1977.

APPENDIX 1

Items in Experiment 1

<u>4-word sentences</u>

1. Touch the red circle.
2. Touch the yellow square.
3. Touch the white circle.
4. Touch each yellow one.
5. Touch the squares quickly.
6. Touch both green ones.

<u>5-word sentences</u>

1. Touch the large green circle
2. Touch the small green circle.
3. Touch the small blue circle.
4. Touch only two big ones.
5. Touch one that isn't white.
6. Touch two that are yellow.

<u>8-word sentences</u>

1. Touch the yellow circle and the red square.
2. Touch the green square and the blue circle.
3. Touch the blue square and the yellow square.
4. Touch the blue circle with the red square.
5. Touch the blue circle or the red square.*
6. Except for the green one, touch the circles.

<u>10-word sentences</u>

1. Touch the small yellow circle and the large green square.
2. Touch the small blue square and the small yellow circle.
3. Touch the small blue circle and the small red circle.
4. Pick up all the circles except for the yellow one.*
5. Before you touch the green square, touch the white circle.*
6. If there is a yellow square, touch the green circle.*

* = modified Token Test item

DISCUSSION

C: The way you talk about your two variables I think illustrates some of the needs we have for theory from people working on what the cognitive processor is, because it is not unusual to equate the variable of length to processing, as at least you do in your abstract. And [you imply] that linguistic complexity is something else; that it's a syntactic variable, it's a grammatical variable that somehow is not linked explicitly to processing. It seems to me that we need to try to characterize each of these variables with respect to the kind of processing each one relates to rather than treat length as related to processing, and this grammatical variable as something else. I think your linguistic analyses are ways of approaching the requirements on the processor that might lead to then saying one variable deals with short-term memory, and another variable, the complexity variable, deals with some other aspect of the processing mechanism.

A: That's a good point, and when we were thinking about this we were very much aware that it was beginning to sound like processing had nothing to do with grammar, and I didn't mean to imply that. Certainly there's syntactic or grammatical processing, and it probably would be more accurate to get rid of the word "processing," at least in terms of contrasting short-term memory and grammar, because both involve processing.

C: I think I'm a little more bothered by the term memory. I think that processing is appropriate, but it might be a bit of an inferential leap to say that length is just memory.