Effects of Syntactic Confusion on Semantic Comprehension in Aphasia

Edythe A. Strand and Malcolm R. McNeil
University of Wisconsin-Madison, Madison, Wisconsin

There has been a great deal of interest in examining auditory language comprehension in terms of the relative contribution of lexical/semantic and syntactic processes. Investigators have used semantic or syntactic tasks to differentiate aphasic from nonaphasic persons (Pizzamiglio and Appicciafuoco, 1971; Lesser, 1974; Orgass and Poeck, 1966), as well as to provide evidence for differences among the different aphasic syndromes (Shevan and Canter, 1971; Von Stockert, 1972; Zurif, Caramazza, and Myerson, 1972; Zurif, Green, Caramazza, and Goodenough, 1976; Caramazza and Zurif, 1976; Von Stockert and Baker, 1976; Goodglass, Blumstein, Gleason, Hyde, Green and Statlender, 1979).

It is commonly accepted that semantic and syntactic processes interact in auditory comprehension. There remains, however, some controversy regarding whether there are differences in the relative contributions of semantic and syntactic processes to the auditory comprehension deficits exhibited by the different aphasic subgroups (Shevan and Canter, 1971; Orgass and Poeck, 1966; Poeck, Kerchensteiner, and Hartje, 1972; Odell, 1983). Similarly, the degree to which tests of auditory comprehension of sentences evaluate specific deficits in each of these processes is still unclear.

The Token Test, which is widely used for examining auditory comprehension of sentences, is purported to be especially sensitive to aphasic behaviors (Orgass and Poeck, 1966). It is not yet clear, however, which of these linguistic processes contribute to that sensitivity. Poeck et al. (1972) found that auditory comprehension deficits, as measured by the Token Test, are common to different clinical types of aphasia. They compared fluent and nonfluent aphasic subjects on the Token Test and found no significant difference between distributions of low and high scores between the two groups. Comparison between the mean scores of the two groups for each part of the test also yielded no significant difference. Their data indicated, therefore, that the Token Test could not be used to differentiate between clinical types of aphasia.

The finding that there is no quantitative difference in language understanding as measured by the Token Test is contradictory to the common view that there exist qualitative and quantitative differences in auditory comprehension among various types of aphasia. The findings reported by Poeck et al. (1972), however, speak only to overall performance on the Token Test. The data do not address specific factors underlying that performance.

It was suggested by DeRenzi and Vignolo (1962), the creators of the Token Test concept, that the difficulty in performing Token Test commands results from the necessity to grasp the meaning of each word in the command, because there are no predictable cues from syntactic form. Recently it has been suggested that there are syntactic effects which do affect performance on the Token Test (Vermeulen, 1982). Vermeulen reported the results of a study on a large mixed sample of aphasic subjects without regard to clinical type, in order to demonstrate that levels of remaining function in the areas of lexical/semantic and syntactic comprehension can vary independently of one another. He administered nine auditory comprehension tests specifically aimed at either semantic or syntactic comprehension. Through the use of

factor analysis, he found that even in a general population of aphasic patients, separation of two auditory comprehension factors can be delineated, which he interpreted as lexical comprehension and syntactic comprehension.

Vermeulen (1982) also examined the relationship of the resulting factors to the standardized Dutch version of the Token Test (Van Dongen, Van Harskamp, Verhey-Stoll and Lutfijn, 1974). The results indicated that only the subgroup with high scores on both factors approached a normal level of Token Test performance. Similar results were obtained for each part of the Token Test analyzed separately. He concluded that disruption of either a semantic or a syntactic factor seems likely to impair overall performance.

One interesting finding noted by Vermeulen was that even in Part 1 of the Token Test (e.g., "Touch the red circle") syntactic comprehension seemed to be involved. He argued that in order to execute this command, a connection needs to be made between the noun and adjective concepts. Vermeulen noted that, while this would seem possible without syntactic information, aphasic subjects apparently have to use syntactic processing when attempting the task.

The present study was designed to provide data to confirm or refute Vermeulen's interpretation relative to the syntactic load on the Token Test. In order to manipulate the syntactic contribution to the task, six different word orders of the Revised Token Test (McNeil and Prescott, 1978) sentences were constructed. The manipulation involved permuting the size, color and shape elements into all possible orders. The experimental question addressed in this study was: Are there significant differences in performance of RTT sentence commands across different word order conditions for aphasic or for normal individuals? No differences across conditions would support a syntactic insensitivity or semantic/lexical sensitivity for the task. Differences in responses between normal English word order and permuted word orders would be consistent with syntactic sensitivity for the tasks.

METHOD

Subjects. Ten aphasic subjects, each with a left hemisphere lesion due to cerebral vascular accident served as the experimental group. Each subject achieved either a 30th percentile score or higher on the 5-item Revised Token Test (RTT) (Arvedson, McNeil, and West, 1985), or a 40th percentile score or better on the shortened version of the 2-item-per-subtest Porch Index of Communicative Ability (SPICA) (DiSimoni, Keith, Holt, and Darley, 1980). The aphasic subjects ranged in age from 48-71 with a mean age of 53.8. All subjects were native speakers of English, had had some period of speech/language therapy following their stroke and had passed audiological screening at 35 dB at 500, 1K, and 2K. A battery of linguistic tests were administered in order to provide descriptive data regarding the speech and language performance of each subject. This battery included the auditory comprehension subtests of the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass and Kaplan, 1983), the Auditory Comprehension Test for Sentences (ACTS) (Shewan, 1979), and a spontaneous language sample from which further observations were made regarding the patient's speech and language (Table 1).

Five age-matched adults who reported no history of neurological problems and who exhibited no speech, motor or cognitive problems also served as subjects. The normal group, who ranged in age from 56 to 70, were given the same battery of speech and language tests as the aphasic group. All scores on language testing were within normal limits as measured by the norms for each measure. Differences in overall performance between the two groups is

to be expected and was not the question of interest. The normal group was included in this study in order to examine possible qualitative differences in patterns of performance when confronted with word order confusion.

Table 1. Descriptive data and summary of test results for the ten aphasic subjects.

	Sex	Age	PICA %ile	RTT %ile	ACTS	BDAE
S1	M	71	34	49	19	79
S2	M	66	45	32	16	101
S3	M	59	56	39	17	114
S4	M	48	73	22	15	109
S5	M	55	75	35	17	103
S6	M	59	75	52	17	112
S7	M	66	79	16	15	102
S8	M	61	38	32	8	118
S9	M	66	78	75	20	115
S10	M	60	90	74	20	107

Procedures. Ninety test sentences, comprised of six word order conditions, were derived from the first five sentences of subtests II, IV, and VI of the RTT (Table 2). These subtests were chosen so that the data relative to word order differences could also be examined with respect to factors of length and the addition of a prepositional element.

Table 2. Examples of RTT subtests II, IV and VI.

II. Touch the big green circle.
IV. Touch the little blue square and the big black square.
VI. Put the big red square in front of the big white circle.

Each of the five sentences were permuted six ways in order to allow all possible orders of size, color and shape (Table 3). These permutations yielded 30 stimulus sentences for each of the three subtests. The six conditions were arranged randomly for each subtest, and presented in that same random order for each subject. Order of presentation of subtests II, IV, or VI was random for each subject.

Table 3. Examples of the six word order conditions.

1. Touch the big green circle.
2. Touch the circle big green.
3. Touch the green circle big.
4. Touch the circle green big.
5. Touch the big circle green.
6. Touch the green big circle.

Prior to administration of the experimental stimuli, each subject was presented with the RTT tokens and screened to verify their ability to make consistent correct pointing responses to particular shapes and colors. The stimulus sentences were presented free-field via cassette audio tape. The subjects were instructed to listen carefully to the sentences on the tape and to do as each sentence commanded. Each subject was allowed two repetitions of each command, which were given following either a request for a repetition, no response, or if the subject did the wrong task.

The multidimensional scoring system for the RTT (McNeil and Prescott, 1978) was used to record the subject's response for each element of each test sentence. The average of the total score for each sentence was computed. The measure used for analysis was the average of the five sentence means for each of the six word order conditions for each subject. A univariate repeated measures ANOVA was used to analyze the collected data (Davidson and Toporek, 1981).

RESULTS AND DISCUSSION

A significant difference was found in performance for the main effect of subtest ($p < .05$) in the aphasic group (Figure 1). Scheffe post-hoc comparisons revealed differences among all three subtests. In the normal group, no difference was found between subtests II and IV or between subtests IV and VI. A difference was found between subtests II and VI. While this finding was not unexpected, it indicates that the addition of both the factors of increased length and a prepositional element did decrease performance in the normal group (Figure 1).

No significant difference ($p > ,05$) was found among the six word order conditions in each subtest for either the aphasic or the normal group. Figure 2 depicts the data for the aphasic group. All aphasic subjects performed as the group did in that no subject performed consistently better or worse with any particular word order.

The results also indicated no significant interaction for subtest and condition in either group of subjects. It seems, therefore, that the differences found between the subtests were not affected by permuting the word order of the sentences. Subjects performed similarly across all six word order conditions in each of the subtests. Neither additional sentence length nor addition of a prepositional element caused a difference in performance across different word orders. It is interesting to consider why syntactic anomalies did not affect subtests differentially. While it has been suggested that nonlinguistic cognitive factors such as memory are

important in completing Token Test tasks (Lesser, 1976), it is not clear which cognitive operations may be independent from linguistic processes. While the data from this study do not directly address this issue, it is clear that length was a more differentiating factor than word order in auditory comprehension of these simple commands and that the effects of permuted word order did not differ across sentences of different lengths.

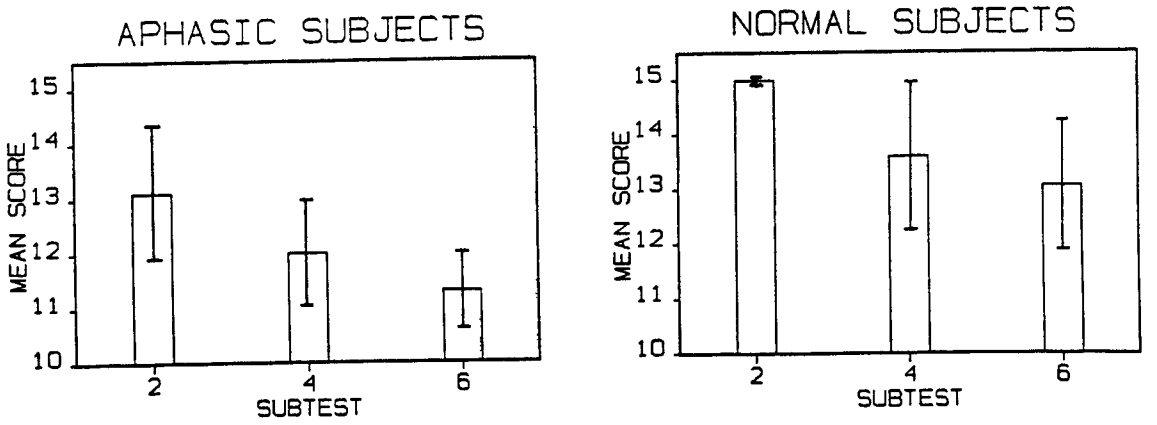

Figure 1. Aphasic (L) and Nonaphasic (R) group performance for subtests II, IV, and VI of the experimental RTT task. Means are derived from the 30 stimulus items across the six word order conditions for each of the subtests.

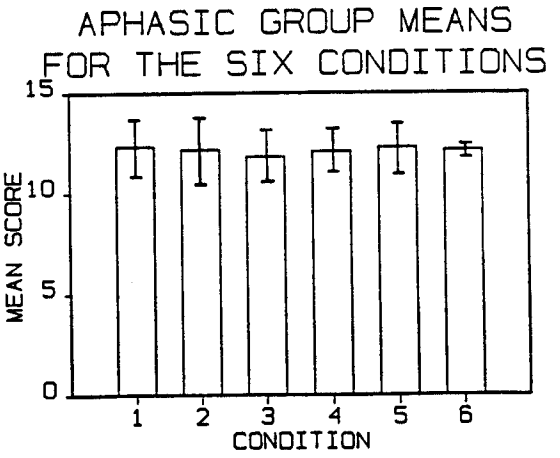

Figure 2. Aphasic group means and standard deviations for each of the six word order conditions across subtests II, IV and VI of the experimental RTT task.

The finding of no difference among the six word order conditions may be interpreted as evidence that the RTT is a lexically or semantically weighted test which is not sensitive to syntactic factors that may be operating in other auditory comprehension contexts. These data seem inconsistent with Vermeulen's (1982) conclusion that syntax is an important factor even for the simplest commands of the Token Test. It should be noted, however, that our data reflect only adjective/noun word order confusion. Other syntactic elements such as noun-verb relationships may be important to his factor analysis findings and should be investigated as to their importance in performance on these tasks.

The finding of no difference among correct word order and five incorrect orders in the aphasic group may be examined from a variety of perspectives. If one accepts these data to suggest that semantic cues are salient on the Token Test, and if one accepts that semantic and syntactic systems are differentially impaired in aphasic subjects (Von Stockert, 1972; Zurif, Caramazza, and Merson, 1972; Lesa, 1974; Zurif, Green, Caramazza, and Goodenough, 1976; Goodglass, Blumstein, Gleason, Hyde, Green, and Statlender, 1979), it would seem logical that differences in performance on the Token Test would be predicted for the Wernicke and Broca classifications. Yet, the opposite has been demonstrated (Poeck et al., 1972).

Studies investigating auditory comprehension of whole sentences provide data that are relevant to these seemingly inconsistent findings. Gardner (1975) found that all aphasic groups exhibited semantic errors. While his study did not examine syntactic and semantic comprehension differences, his data point out that both Wernicke and Broca aphasic subjects exhibit semantic deficits in comprehension of spoken sentences. Shewan and Canter (1971) also examined auditory comprehension of whole sentences. They attempted to determine to what extent various aphasic subgroups would be differentially affected by changes in syntactic structure, vocabulary difficulty and sentence length. While anomic, Broca and Wernicke aphasic groups displayed different levels of impairment (Wernicke greater than Broca, and Broca greater than anomic), the effect of increasing the difficulty of each of the three factors was the same for all three groups of aphasic subjects. The authors concluded that there is no qualitative difference in the type of comprehension problems exhibited by the different aphasic groups. Given these data indicating that both Broca and Wernicke patients exhibit difficulty with auditory comprehension of sentences, it is not surprising that the Token Test, even if semantically weighted (as our data suggest), does not differentiate among aphasia classifications. The Revised Token Test may examine performance at a very basic level of language processing that taps performance deficits common to all aphasic patients.

During administration of the experimental stimuli, it was often noted that subjects seemed to employ specific strategies for processing the auditory information and completing the command. For example, certain "trade offs" seemed to be made (i.e., tuning into size and ignoring color; or always using delays, or always using immediacy). By stressing the system through mixing up word order, these strategies seemed even more obvious. One aspect to consider concerning this subjective observation is that of perceptual salience. The aphasic patient will sometimes pay more attention to one particular perceptual variable. This behavior is evident in a task such as the Token Test where aphasic subjects might reduce attention to one or more elements in an effort to respond correctly to other elements in the sentence. Kriendler et al. (1971) presented data indicating that errors for shape were not due to serial position in the sentence. If subjects were attending to a particular item type in the task in this study, then the finding of no differences across word order conditions would not be surprising. Subjects would be able to attend to a particular item or two to the exclusion of the others regardless of the word order in which it was presented.

In conclusion, we suggest that our data lend support to the hypothesis that the Revised Token Test is semantically weighted and that it may test primarily semantic processes in auditory comprehension. Further research might profitably be directed toward those cognitive operations that support and interact with syntax and semantics.

REFERENCES

Arvedsen, J.C., McNeil, M.R. and West, T.L. Prediction of Revised Token Test overall, subtest, and linguistic unit scores by two shortened versions. In R.H. Brookshire (Ed.), Clinical Aphasiology: Conference Proceedings, 1985. Minneapolis, MN: BRK Publishers, 57-65, 1985.

Caramazza, A. and Zurif, E. Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. Brain and Language, 3, 572-582, 1976.

Davidson, M. and Toporek, J. General univariate and multiple analysis of variance and covariance, including repeated measures. In W. Dixon (Ed.), BMDP Statistical Software. Berkeley, CA: University of California Press, 1981.

DeRenzi, E. and Vignolo, L. The Token Test: A sensitive test to detect receptive disturbances in aphasics. Brain, 85, 556, 1962.

DiSimoni, F.G., Keith, R.L., Holt, D.C. and Darley, F.L. Prediction of PICA overall score by short versions of the test. Journal of Speech and Hearing Research, 23, 511-516, 1980.

Gardner, H., Albert, M., and Weintraub, S. Comprehending a word: The influence of speed and redundancy on auditory comprehension in aphasia. Cortex, 11, 155-162, 1975.

Goodglass, H., Blumstein, S., Gleason, J., Hyde, M., Green, E. and Statlender, S. The effects of syntactic encoding on sentence comprehension in aphasia. Brain and Language, 7, 201-209, 1979.

Goodglass, H. and Kaplan, E. The Assessment of Aphasia and Related Disorders. Philadelphia, PA: Lea and Febiger, 1983.

Kreindler, A., Gheorghita, N., and Voinescu, I. Analysis of verbal reception of a complex order with three elements in aphasics. Brain, 94, 375-386, 1971.

Lesser, R. Verbal comprehension in aphasia: An English version of three Italian tests. Cortex, 10, 238-246, 1974.

Lesser, R. Verbal and non-verbal memory components in the Token Test. Neuropsychologia, 14, 79-85, 1976.

McNeil, M.R. and Prescott, T.E. Revised Token Test. Baltimore, MD: University Park Press, 1978.

Odell, K.A. Comparisons between auditory and reading comprehension in aphasic adults. Masters Thesis, University of Wisconsin, Madison, Wisconsin, 1983.

Orgrass, B. and Poeck, K. Clinical validation of a new test for aphasia: An experimental study on the Token Test. Cortex, 2, 224-243, 1966.

Pizzamiglio, L. and Appicciafuoco, A. Semantic comprehension in aphasia. Journal of Communication Disorders, 3, 280-288, 1971.

Poeck, K., Kerschensteiner, M. and Hartje, W. A quantitative study of language understanding in fluent and nonfluent aphasia. Cortex, 8, 299-304, 1972.

Porch, B. Porch Index of Communicative Ability. Vol. 2: Administration, Scoring and Interpretation. Palo Alto, CA: Consulting Psychologists Press, 1971.

Shevan, C. and Canter, G. Effects of vocabulary, syntax and sentence length on auditory comprehension in aphasic patients. Cortex, 7, 209-226, 1971.

Shevan, C.M. Auditory Comprehension Test for Sentences. Chicago, IL: Biolinguistics Clinical Institutes, 1979.

Van Dongen, H., Van Harskamp, K., Verhey-Stoll, F., and Lutfijn, F. Afasie-

onderzoek met de Token Test. Enige psychometrische kenmarken en evaluatie van een verkorte vorm. Nederland Tijdschrift voor de Psychologie, 28, 633-647, 1974.

Vermeulen, J. Auditory language comprehension in aphasia: A factor-analytic study. Cortex, 8, 287-300, 1982.

Von Stockert, T. Recognition of syntactic structure in aphasic patients. Cortex, 8, 323-334, 1972.

Von Stockert, T.R. and Baker, L. Some relations of grammar and lexicon in aphasia. Cortex, 12, 49-60, 1976.

Zurif, E., Caramazza, A., and Myerson, R. Grammatical judgements of agrammatic aphasics. Neuropsychologia, 10, 405-417, 1972.

Zurif, E., Green, E., Caramazza, A., and Goodenough, C. Grammatical intuitions of aphasic patients: Sensitivity to functors. Cortex, 12, 183-186, 1976.

DISCUSSION

Q: When you ask a person to touch the circle big green, the person has to identify the circle; they have to touch the circle, not something big or green.

A: That's right.

Q: It seems to me that you're attributing a semantic processing characteristic to the stimulus. Might it not be possible that the person would use their syntactic knowledge to rearrange the items in a command?

A: The subjects didn't seem to be reorganizing word order in order to cue into a particular semantic label in order to get the task right. Because we scored every single element and then looked at all those elements individually as well as across the means, it didn't matter where the element was in the serial position of the sentence in terms of what they were cueing into.

Q: This study seems to speak to the resistance of this sort of lexical, inferential route to aspects of meaning in the face of syntactic deformation. I just wonder how resistant it is. So, for instance, in the latter sections of the Token Test, I think you could deform the sentences along lines like touch the circle white and black square brown red. (There should be a "the" in there somewhere.) And given the nature of the tokens there is only one solution to the problem. However, in order to infer that solution from lexical content, you would have to recognize that the adjectives were placed outside of their noun phrases in another conjunct. And you'd have to totally ignore the syntax. And I wonder whether you'd get equally good performance if you made those kinds of larger syntactic deformations within the noun phrases themselves.

A: That's a good question and it's something that we've talked about both before and since doing this study. By choosing only subtests II, IV and VI we're very limited in the conclusions we can draw about word order confusion. These conclusions have to be limited strictly to adjective-noun contexts without taking into account what kind of syntactic processing is occurring in terms of the conjunctions that are important in those particular tasks. The study that we reported today can only be examined in light of subtests II, IV and VI. It would be very interesting to see what would happen if we took the more complex tasks and mixed up even prepositional elements.

Q: I'm curious about subjects' use of strategies and how they emerged when you stressed the system. Could you comment about the consistency with which individual subjects would use a strategy across stimuli?

A: I do have some data with respect to how they used some strategies such as immediacy, delay and vocalizations and how those strategies varied across subtest. There was little variation across conditions. The length of the stimuli seemed to contribute more to the increased use of these strategies than word order. Even within a particular subject, their use of these strategies increases with increased length of stimuli. For subtest II, 9 of the 10 subjects used delay at least some of the time, while none used immediacy and one always used vocalization. For subtests IV and VI, however, 7 of the 10 subjects used immediacy as well as delays in trying to respond correctly.

Q: Otherwise, individual subjects did tend to switch strategies across subtests?

A: Yes. I expected different subjects to use different strategies in general, but I was surprised to note that as the length increased individual subjects would switch or use additional strategies.

Q: Regardless of the tasks, it seems they were challenged by the situation of dealing with funny atypical word orders. I was wondering if you have any evidence that that skill in dealing with that constant problem changed over time.

A: I did not examine that statistically, but I did pay attention to that issue as I was computing all the means for analysis. My subjective impression is that they did not do any better in the later items compared with the first items in each subtest.