Brain and Language **72,** 193–218 (2000)
doi:10.1006/brln.1999.2285, available online at http://www.idealibrary.com on IDEAL®

Quantitative Analysis of Aphasic Sentence Production: Further Development and New Data

Elizabeth Rochon,*,† Eleanor M. Saffran,†,‡ Rita Sloan Berndt,§ and Myrna F. Schwartz†

**University of Toronto, Toronto, Canada; †Moss Rehabilitation Research Institute, Philadelphia, Pennsylvania;‡ Temple University, Philadelphia, Pennsylvania; and §University of Maryland School of Medicine, Baltimore, Maryland*

The narrative production of patients with Broca's aphasia and age-and education-matched control subjects was analyzed using the Quantitative Production Analysis (Saffran *et al.*, 1989), a procedure designed to provide measures of morphological and structural characteristics of aphasic production. In addition to providing data for a larger number of subjects than in the original study, we provide data on interrater and test-retest reliability. The data were also submitted to factor and cluster analyses. Two factors characterized the data and the cluster analysis yielded four sets of patients who performed differently on these factors. In particular, there is evidence that agrammatic patients can differ in their production of free and bound grammatical morphemes, substantiating earlier claims in the literature.

Key Words: aphasia; agrammatism; sentence production; syntax; morphology.

Saffran *et al.* (1989) developed the Quantitative Production Analysis (QPA) as a means of quantifying syntactic aspects of aphasic production. In particular, it was designed to capture and describe the speech production patterns of patients labeled ''agrammatic,'' and as such places most emphasis on grammatical structures. The QPA is performed on samples of narrative speech, elicited by asking patients to relate a well-known fairy tale or other familiar story, usually ''Cinderella.'' In developing the QPA, the authors departed from the usual error-based approach to the analysis of aphasic production. As the nature of the error can be difficult to specify in contexts that provide little syntactic information, the majority of the QPA measures focus, instead, on the frequency of occurrence of features such as grammatical morphemes, propositional utterances, and the extent to which utterances are elaborated beyond the minimal requirements for a ''sentence'' (noun plus verb).

The purpose of the analysis system is to provide both an objective means of comparing deficits across patients, as well as to detect changes that occur

0093-934X/00 $35.00

in a single patient's production impairment over time. The QPA can be used to identify patients for research, to assist in the differential diagnosis of non-fluent aphasic patients, or to measure changes in production with recovery or following treatment. There are numerous examples of studies that have utilized the QPA successfully for these purposes (e.g., Byng, 1988; Byng *et al.,* 1994; Martin *et al.,* 1989; Schwartz *et al.,* 1994; Hesketh & Bishop, 1996; Bird & Franklin, 1996; Edwards, 1995).

In addition to providing instructions for performing the analysis, the 1989 paper reported QPA data for a small number of patients ($N = 10$) and normal controls ($N = 5$). The patients were Broca's aphasics, half of whom were clinically identified as agrammatic. Agrammatic and nonagrammatic patients differed from the controls with respect to speech rate (words per minute). Both sets of patients showed reduced sentence complexity relative to the controls, but only the agrammatics deviated significantly from normals on indices of morphology, such as percentage of closed class words and the use of inflections. Nevertheless, the distribution of scores on morphological measures suggested continuity between the two groups, as opposed to a clear distinction. Others have also recently reported continuity across types of aphasia (e.g., from agrammatic, through nonfluent nonagrammatic, through fluent patients), on sentence complexity and grammatical morpheme indices of the QPA (Hesketh & Bishop, 1996). Inspection of the individual patient data in Saffran *et al.* (1989) also pointed to different patterns of morphological breakdown within the agrammatic group: e.g., one patient demonstrated good control of bound morphemes although he produced few free-standing grammatical morphemes, while another exhibited the opposite pattern. Other investigators have reported similar variability among agrammatic speakers (e.g., Miceli *et al.,* 1989; Bird & Franklin, 1996).

The restricted size of the data set was a distinct limitation of the Saffran *et al.* (1989) study. We have since obtained data from 29 patients classified as Broca's aphasics, and 12 age-matched controls. The purpose of the present study was to present more data, based on a larger sample size, and, with this large sample, to further explore the findings from Saffran *et al.* (1989). In particular, we were interested in the dissociation in the production of bound and free-standing grammatical morphemes that was noted in the agrammatic group and in the finding that nonfluent, nonagrammatic patients also showed a reduction on measures of sentence complexity.

The present investigation also takes a different approach to the analysis of the QPA data. Instead of classifying the patients a priori as agrammatic or nonagrammatic, and comparing groups so defined, the data were subjected to factor and cluster analyses to determine (1) whether particular measures would pattern together and (2) whether different performance patterns would emerge within the patient group. We also report reliability measures, based on data from the new, larger sample. These data should prove useful to those who choose to implement the QPA in their studies. Our findings also provide further definition of the sentence production deficits of patients classified as Broca's aphasics.

METHODS

Subjects

A new sample of 29 chronic aphasic patients participated in this study. Subjects were referred to the study from laboratories in Philadelphia, Baltimore, and Houston. All patients had suffered a single, left-hemisphere cerebrovascular accident no less than 6 months prior to the initiation of the study. Twenty-five subjects were right-handed and four were left-handed. There were 19 men and 10 women. Patients ranged in age from 22 to 74, with a mean age of 54.4 (SD = 12.0). They had from 11 to 20 years of education with a mean of 14.1 years (SD = 2.5). All subjects were native speakers of English.

All patients were classified as Broca's aphasics on the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983). They demonstrated good auditory comprehension on the BDAE and produced sparse, halting speech, with some apparent effort. Twenty patients were classified as agrammatic by the referring speech-language pathologist or neuropsychologist; nine were classified as nonfluent nonagrammatic patients.

Twelve control subjects also participated, matched as closely as possible to the patients on the basis of age and education. Eleven were female and one male. Control subjects were aged 20 to 73 with a mean of 51.3 years (SD = 13.6). They had from 12 to 16 years of education with a mean of 13.75 years (SD = 1.86).

Procedures

The procedures outlined in Saffran *et al.* (1989)[1] were employed for eliciting and transcribing the speech sample, as well as for extracting the narrative word corpus and segmenting it into utterances for analysis. More detailed instructions for using the Quantitative Production Analysis (QPA) can be found in Berndt *et al.* (2000). In this paper, further reference to the measures used in the QPA will be to Berndt *et al.* (2000). Briefly, narratives were elicited by having patients produce a fairy tale, usually ''Cinderella,'' which was tape recorded. If the Cinderella narrative yielded fewer than 150 narrative words, the subject was asked to tell another well known

[1] Calculation of two measures for this study was modified from the original method of calculation in Saffran *et al.* (1989). These were the proportion of pronouns to nouns (P/N+P) and the proportion of verbs to nouns (V/N+V). Two other measures were included in this study that were not in the original study. These were the measure of struggle that was based upon the proportion of narrative words to total words uttered in a narrative sample and the median length of utterance.

or familiar story (e.g., ''Little Red Riding Hood,'' ''Jack and the Beanstalk,'' etc.) in addition to ''Cinderella.'' After transcription, the speech sample was timed for speech rate in words per minute (WPM). Next, a corpus of narrative words was extracted. Procedures for extracting narrative words were designed to isolate the propositional speech produced by the patient from other frequent manifestations of nonfluent aphasic speech, such as false starts, repeated attempts at a word, stereotyped utterances, perseverations, responses to the examiner's questions, etc. The QPA was carried out on each subject's first 150 narrative words (plus or minus 10). This ''narrative core'' was segmented into utterances, using prosodic features as well as syntactic structure to determine utterance boundaries, and analyzed for lexical, morphological, and structural measures.

Subjects' production of lexical items of various types (see below) was calculated independently of utterance type across the entire narrative sample. Though not exhaustive, this tabulation is intended to capture important aspects of lexical and structural abnormalities found in nonfluent aphasic speech (Saffran *et al.*, 1989; Berndt *et al.*, 2000). Numbers of narrative words, open class words, nouns, determiners, pronouns and verbs were counted. Also enumerated were the numbers of nouns requiring determiners, verbs that could be inflected given the context, and verbs that did carry an inflection. The following measures were generated on the basis of these counts of lexical items: proportion of narrative words that were closed class (CC); proportion of determiners produced in obligatory contexts (Determiner index); (personal) pronouns as a proportion of (personal) pronouns + nouns (P/N+P); verbs as a proportion of verbs + nouns (V/N+V); proportion of verbs inflected in contexts that would permit inflection (Inflection index); speech rate (WPM)[2]; proportion of narrative to total words uttered in the narrative sample (Struggle measure).

Utterances that were designated minimal ''sentences'' (defined as a noun + a verb) were subjected to further analyses designed to assess the morphological complexity of the verb phrase and the structural complexity of sentential elements (NP, VP). The latter, based on a count of open class words, prepositions and pronouns, yields a general index of the extent to which sentences are elaborated beyond their minimal (N + V) constituents. The number of embeddings within each sentence was also calculated. The following measures were produced as a result of sentence-based calculations: an index of elaboration of the auxiliary in matrix verbs (AUX score); proportion of narrative words occurring in sentences (Proportion words in sentences); proportion of sentences that were well-formed (Proportion well-formed sentences);[3] an index of the combined structural elaboration of NP and VP sen-

[2] Words per minute is the only measure that is calculated on the entire speech sample rather than the 150-word narrative sample that is used to calculate all other measures.

[3] To count as well-formed, a sentence had to be syntactically well-formed but could be semantically anomalous (e.g., ''he was the bean.''). This and other scoring conventions are described in Berndt *et al.* (2000).

tence constituents (Sentence elaboration index); number of embedded clauses per sentence (Embedding index); median length of utterance (MedLu) was also calculated based upon all utterances.

RESULTS AND DISCUSSION

Source of Narrative Samples

To extract 150 word narrative samples, patients told a mean number of 2.2 (SD = 1.2) stories. For all control subjects, only one story (in all cases, ‘‘Cinderella’’) was required to elicit the requisite number of narrative words.

Reliability Measures

Reliability was assessed in two different ways. (1) Interrater reliability, which involved the analysis of the same set of speech samples by two different scorers; this measure examines consistency across raters. (2) Test-retest reliability, where the same scorer analyzed two different samples, obtained at different times, from the same subject; this measure examines consistency of the scores across speech samples.

(1). Interrater Reliability

The starting point for the analysis system is the set of narrative words that remains after false starts, perseverations, and extraneous comments on the narrative have been removed. Detailed criteria have been devised for the extraction of this narrative core, and the extent to which they are applied consistently by different examiners was investigated. Next, an analysis was carried out on the specific scoring decisions made for each measure. Eight scorers from three separate sites participated in scoring the narratives. These included three of the authors and five other scorers, three of whom were speech-language pathologists and two of whom were research assistants. Scorers unfamiliar with the analysis system first practiced segmenting and scoring transcripts from the corpus of transcripts available from Saffran *et al.* (1989). Difficulties and discrepancies were discussed with and resolved by one of the first three authors. Efforts were made to achieve comparability of scoring ability across scorers and across the two populations (aphasic patients and control subjects). However, since the QPA was designed for use with aphasic patients, scorers received more practice with aphasic transcripts and less with control transcripts. To examine reliability, the results obtained by two independent scorers were compared in two separate steps: one for utterance boundary assignment and the other for scoring of the indices.

(i.) *Reliability of utterance assignment.* Twenty-four percent of the aphasic patients’ narratives ($N = 7$) and 33% ($N = 4$) of the control subjects’ narratives were segmented into utterances by two independent scorers. Six of the eight scorers contributed to this measure. Narratives distributed to scorers

TABLE 1
Interrater Reliability: Proportion of Agreement in Extraction of Narrative Words and Utterance Boundary Assignment for Both Groups

	Patients	Controls
1. Extraction of narrative words		
Mean	.97	.98
SD	(.03)	(.02)
Range	.90–1.0	.96–1.0
2. Assignment of words to utterances		
Mean	.90	.97
SD	(.08)	(.04)
Range	.76–1.0	.93–1.0

were randomly selected. Scorers received an audiotape of a subjects' narrative and a loose transcription of that narrative. Their instructions were to extract narrative words and to segment the transcript into utterances as per the instructions in Saffran *et al.* (1989). The proportion of words that both scorers assigned to the same category (+narrative word or −narrative word) was calculated, as was the proportion of narrative words that both scorers assigned to the same utterance.

Table 1 shows the proportion of narrative words that two scorers assigned to the same category (+narrative word or −narrative word) in the extraction of the 150-word narrative core. As is evident from the table, agreement between two scorers as to what constituted a narrative word was .97 for aphasic patients and .98 for controls. This replicates findings from Saffran *et al.* (1989) for patients and indicates a high degree of agreement between scorers, for both groups of subjects.

As can also be seen in Table 1, the mean proportion of words assigned to the same utterance agreed upon by two scorers was .90 for patients and .97 for controls. It should be noted that though the range for the patient data is quite large, the lowest agreement score, .76, is an outlier. All other scores fell between .85 and 1.0. These findings again replicate results from Saffran *et al.* (1989) for the aphasic patients, and, for both groups, demonstrates substantial agreement as to what constitutes an ''utterance.'' Thus the instructions for utterance boundary segmentation for the quantitative analysis procedure yield highly reliable results, for both aphasic and control narrative samples. This indicates that a narrative core of propositional speech can be reliably extracted using the QPA. The morphological and structural analyses that constitute the bulk of the QPA can then be carried out on this narrative core.

(ii.) *Scoring reliability.* The reliability with which the various indices in the production analysis were scored was also examined. Forty one percent ($N = 12$) of randomly selected aphasic narratives and 66% ($N = 8$) of control

TABLE 2
Interrater Reliability: Intraclass Correlations for Production Analysis Measures

	Patients	Controls
1. Number of closed class words	.98	.85
2. Number of nouns	.97	.80
3. Number of verbs	.96	.62
4. Number of pronouns	.98	.91
5. Number of matrix verbs	.95	.84
6. AUX score	.98	.83
7. Determiner/noun ratio	.98	—[a]
8. Number of ''sentences''	.96	.80
9. Subject NP elaboration	.97	.87
10. VP elaboration	.91	.90
11. Number of embeddings	.89	.65
12. Median length of utterance	.96	.84

[a] Unable to be calculated.

subject narratives were scored by two independent scorers. All eight scorers participated in this analysis. Narratives distributed to scorers were randomly chosen. Scorers received a transcription of the narrative to be analyzed that was already segmented into utterances. This procedure ensured that measurement of scoring reliability between two scorers would be undertaken upon the same words and utterances. Scorers' instructions were to score the segmented transcription as per the instructions in Saffran *et al.* (1989).

Table 2 shows intraclass correlations for agreement between two scorers on indices in the production analysis. Interpretation of the intraclass correlation coefficient is similar to that of a product moment correlation coefficient, but it is a more appropriate measure for examining reliability between two scorers because its calculation depends on the differences between the means and the standard deviations of the two sets of scores (Snedecor & Cochran, 1989). Overall, agreement was quite high, for both aphasics and controls. The percentage discrepancy between scorers was less than 1% (.0075), demonstrating excellent agreement. Because there was very little variability in the data for the determiner/noun ratio in the control subjects, the intraclass correlation for this measure could not be calculated. Only two correlations, both for the control group (the number of verbs (.62) and the number of embeddings (.65)), appear to be markedly lower than correlations for the other indices. It should be noted that the overall percentage discrepancy between scorers was only 7% for the number of verbs counted, which suggests that although this measure may not yield as high a level of agreement as some of the other measures, it can nevertheless be scored reliably.[4] The em-

[4] There were three instances in the control subject transcripts where two scorers counted a slightly different number of utterances in the entire corpus upon which the QPA measures were based. These were errors of addition, which affected the count for the number of verbs more than for other measures.

TABLE 3
Test-Retest: Intraclass Correlations for Agreement by the Same Scorer on Two Different Samples for the Same Subject

	Patients
1. Proportion closed class words (CC)	.85
2. Determiner/noun ratio (Det. index)	.79
3. Proportion of pronouns (P/N+P)	.92
4. Proportion of verbs (V/N+V)	.66
5. Proportion of verbs inflected/possible inflections (Infl. index)	.76
6. Elaboration of auxiliary (Aux score)	.75
7. Proportion of words in sentences	.80
8. Proportion of well formed sentences	.53
9. Structural elaboration of sentences (Sentence elaboration index)	.74

bedding index score (and correlation) most likely reflects scorers' difficulties in scoring complex sentences. In sum, then, both utterance and scoring conventions for the quantitative analysis system can be applied reliably to narrative samples for aphasic subjects in particular, the population for whom the analysis system was developed. Reliability measures were also very good on the whole for the control subjects' transcripts, with the exception of the two measures discussed.

(2.) Test-Retest Reliability

There were eighteen patients from whom narrative samples had been obtained at two separate times. In order to generate two comparable sets of utterances unaffected by any changes in patients' production abilities between samples, two samples were generated for each subject by selecting alternate utterances across the two narratives in a number of ways, so that each sample contained utterances from both test sessions. For example, all the even numbered utterances from the two samples were combined and compared to all the odd numbered utterances from the two samples. The two samples were analyzed by the same scorer and compared for reliability. Table 3 shows intraclass correlations for agreement on production analysis indices from two samples for the same subject. As can be seen in the table, overall agreement is quite high on most measures. However, two correlations, one for the proportion of verbs to nouns in utterances (.66), and the other for the proportion of well-formed sentences (.53), were lower than those for the remainder of the measures. The latter may reflect the relatively small number of sentences produced by some of the subjects.

With respect to the relatively low reliability score for verb production, inspection of the data revealed that only five patients' samples (28%) showed more than a 10% discrepancy across the two samples for this measure. More than half the samples (61%) showed a discrepancy of 6% or less. Several factors appeared to contribute to the disparity for the five patients who showed the most variability in this measure across samples. Fluctuations in noun as well as verb production contributed to the larger discrepancies, but fluctuations in verb production across samples had a greater effect, since the number of verbs produced was typically quite small for each sample. In several cases, inflated verb counts were the result of reduplicative productions that appeared to reflect the patient's desire to emphasize particular points. When repeated words appeared in the corpus, they were retained if they appeared to serve the narrative goal of emphasis. For example, in one sample, one patient produced the utterances ''clacking, clacking, clacking'' to describe Cinderella's carriage ride, and ''hear ye, hear ye'' to signal the search for the owner of the glass slipper. The subject was credited with the production of five verbs for these two minimal propositions (''clacking'' was deemed a verb rather than an adjective based on its dictionary entry). These types of utterances, when they happened to occur in only one of the reliability samples, inflated the number of verbs produced and contributed to the appearance of instability of this measure. Although this factor should be considered in interpreting changes in patients' values on this measure over time, or when comparing data across patients, for the majority of samples these elements do not affect the computation of the verb/noun+verb ratio and cross-sample reliability is acceptable. As mentioned, for all other measures shown in Table 3 test-retest reliabilty is quite high. However, given the difficulties achieving reliability in the two measures discussed above, test-retest reliability of the QPA warrants further investigation.

Production Measures

Group means for production analysis indices for both aphasic and control groups are shown in Table 4. For each measure, results were compared using unpaired *t* tests. As can be seen from the table, all measures yield significant differences between the two groups.

Fig. 1 to 13 display the data on the individual measures for the full set of subjects, patients, and controls. In all cases, there is some degree of overlap in the distributions of the aphasic and control subjects. It is also evident that some of the patients' scores depart markedly from those of controls. The measures that most sharply distinguish patients from controls include speech rate, the determiner index (where omissions by normal subjects were extremely rare), the proportion of words in sentences, the well-formedness measure, the sentence elaboration score, and the median length of utterance.

Factor and Cluster Analyses

Further analyses were undertaken in order to investigate (1) whether any of the QPA variables might form subsets that are relatively independent of one another; and (2) whether separate groups of patients might show different patterns of performance on these variables. To meet these objectives, we

TABLE 4
Group Means and *t* Values for Quantitative Production Analysis Indices

	Patients	Controls	*t* (39)
1. Proportion closed class words/narrative words (CC)			
Mean	.41	.54	3.82***
SD	(.11)	(.04)	
Range	.12–.58	.47–.61	
2. Determiner/noun ratio (Determiner index)			
Mean	.65	.99	4.12***
SD	(.28)	(.02)	
Range	.02–.95	.94–1	
3. Proportion pronouns (P/N+P)			
Mean	.25	.41	2.75**
SD	(.20)	(.09)	
Range	0–.87	.29–.55	
4. Proportion of verbs (V/N+V)			
Mean	.37	.48	3.34**
SD	(.10)	(.06)	
Range	.14–.56	.35–.63	
5. Proportion of verbs inflected/possible inflections (Inflection index)			
Mean	.56	.92	4.19***
SD	(.29)	(.15)	
Range	.05–1	.53–1	
6. Elaboration of auxiliary (Aux score)			
Mean	.78	1.26	4.46****
SD	(.33)	(.26)	
Range	0–1.33	.8–1.71	
7. Proportion of words in sentences			
Mean	.59	.98	5.44****
SD	(.25)	(.05)	
Range	.08–1	.84–1	
8. Proportion of well-formed sentences			
Mean	.56	.95	6.77****
SD	(.19)	(.08)	
Range	0–.82	.75–1	
9. Structural elaboration of sentences (sentence elaboration index)			
Mean	1.42	3.06	9.4****
SD	(.45)	(.64)	
Range	.63–2.37	2.14–4.06	
10. Embedding index			
Mean	.04	.29	7.0****
SD	(.07)	(.17)	
Range	0–.28	0–.50	
11. Median length of utterance (MedLu)			
Mean	3.57	8.17	9.71****
SD	(1.37)	(1.39)	
Range	1.0–7.0	6.5–10.5	
12. Speech rate (WPM)			
Mean	39.01	160.82	13.64****
SD	(19.63)	(37.00)	
Range	16.58–96.16	107.44–232	
13. Proportion of narrative words/total words (Struggle measure)			
Mean	58.39	84.17	5.58****
SD	(15.19)	(7.35)	
Range	23–84	73–95	

* $p < .05$
** $p < .01$
*** $p < .001$
**** $p < .0001$

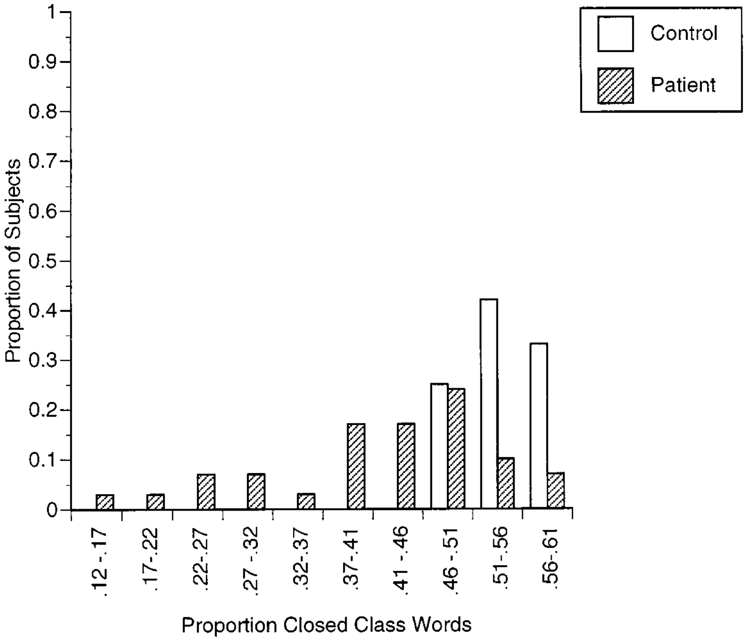

FIG. 1. Frequency distribution of the proportion of closed class words (Number of Narrative Words − Number of Open Class Words/Number of Narrative Words) produced by both groups across the range of scores for that measure.

performed a principal-components analysis (PCA) on the QPA measures to determine whether independent factors composed of sets of correlated variables would emerge from the data. To address the question of whether there were groups of patients that performed differently on these factors, we submitted individual patient factor scores to a cluster analysis. For both these analyses, data from eight patients from Saffran *et al.* (1989) were added to the present corpus, resulting in a sample size of 37.[5] This was done to increase the sample size of the dataset in order that we might submit as many of the QPA variables as possible to the PCA analysis. As a general rule it is preferable to have five cases for each observed variable that is entered into the principle components analysis (Tabachnick & Fidell, 1989). We included nine variables in the PCA, which yielded 4.1 cases per variable. This number is slightly less than the recommended 5; however, it was deemed important to include all relevant variables in order to obtain the most accurate

[5] The eight patients from Saffran *et al.* (1998) were well matched to patients in the present study. Four were diagnosed clinically as agrammatic and four as nonagrammatic. Their mean age was 53.8 and their mean number of years of education was 13.9. New narrative samples had been obtained for the remaining two patients from the 1989 paper. The data for these two patients were already included in the set of 29 patients.

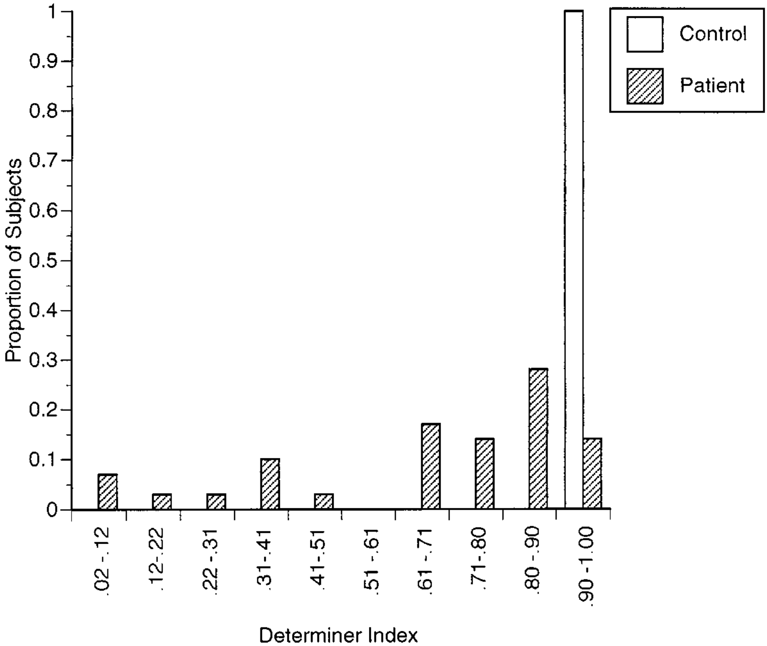

FIG. 2. Frequency distribution of the determiner index score (Number of Nouns Requiring Determiners with Determiners/Number of Nouns Requiring Determiners) for both groups across the range of scores for that measure.

and complete pattern of variable groupings per factor. The results of the PCA are presented below, followed by those of the patient clustering analysis.

Principal-components analysis (PCA). Patient scores on nine of the variables measured in the production analysis were submitted to a PCA. One variable, the proportion of well-formed sentences, was omitted due to the relatively low intraclass correlations obtained when split-half reliability was examined. The embedding index was also removed due to the very small number of embeddings ever produced by patients, and the median length of utterance and the struggle measure were not included as these measures were not available for the eight patients from the 1989 paper (see Footnote 1). It was decided to retain the verb measure despite the somewhat low intraclass correlations obtained with split-half reliability. The inclusion of this variable reflects our assessment of the importance of verb retrieval for the production of sentences to the description and theoretical analysis of agrammatism. The nine variables submitted to the factor analysis included the proportion of closed class words, determiner index, pronouns as a proportion of pronouns + nouns, verbs as a proportion of verbs + nouns, inflection index, auxiliary index, the proportion of words in sentences, structural elaboration of sentences, and words per minute.

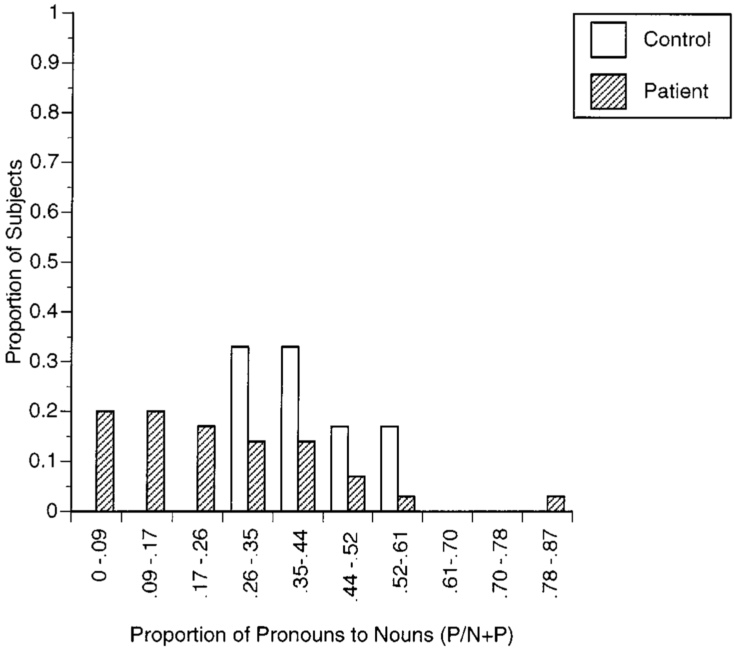

FIG. 3. Frequency distribution of proportion of pronouns (Number of Pronouns/Number of Nouns + Pronouns) for both groups across the range of scores for that measure.

Patient scores on the nine variables were submitted to a factor analysis using a varimax rotation of the principal-components solution. The analysis yielded a 2-factor solution accounting for 70% of the total variance. Table 5 shows the factor loadings. Factor 1 was characterized by high loadings on four variables: the auxiliary index, the inflection index, the proportion of words in sentences, and the structural elaboration of sentences. A second factor was characterized by high loadings for verbs as a proportion of verbs + nouns, the determiner index, proportion closed class words, speech rate, and pronouns as a proportion of pronouns + nouns.

Factor 1, then, appears to be characterized by both bound morphological measures (inflection index, auxiliary index) and structural measures (proportion of words in sentences, the structural elaboration of sentences). Factor 2 is characterized by the relative frequency of verbs to nouns, measures of the frequency of free-standing grammatical morphemes (proportion closed class words, determiner index, proportion of pronouns to nouns), and a fluency, or speech rate measure (words per minute). The variables in Factor 2 appear coherent, in that most reflect the ability to produce free-standing closed class morphemes. It is somewhat surprising that the verb access measure (verb/verb+noun) patterns with these variables, rather than with the structural and aux measures of Factor 1, given that whether an utterance

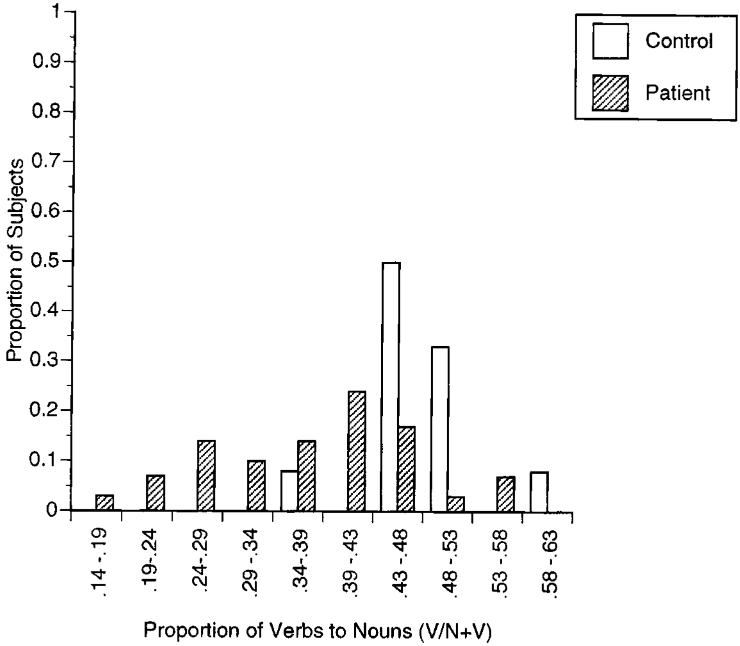

FIG. 4. Frequency distribution of proportion of verbs (Number of Verbs/Number of Nouns + Verbs) for both groups across the range of scores for that measure.

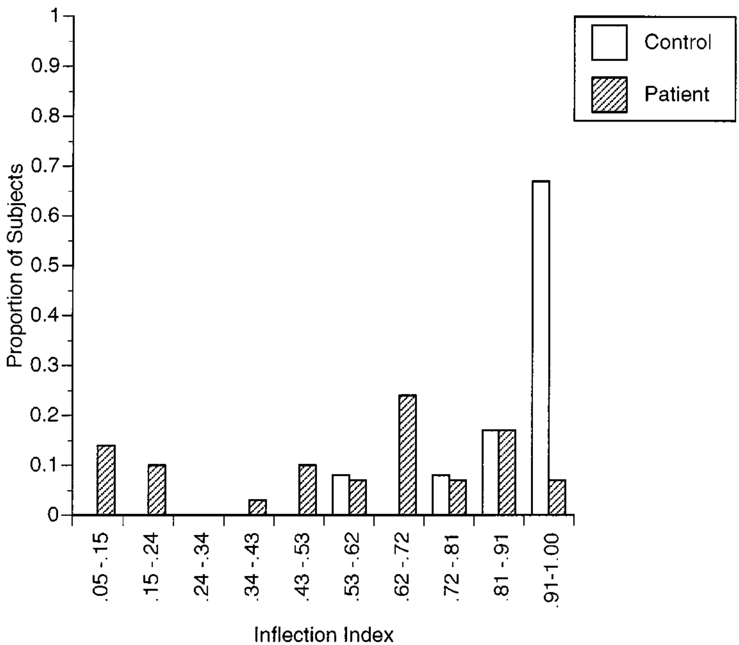

FIG. 5. Frequency distribution of the inflection index score (Number of Inflectable Verbs Inflected/Number of Inflectable Verbs) for both groups across the range of scores for that measure.

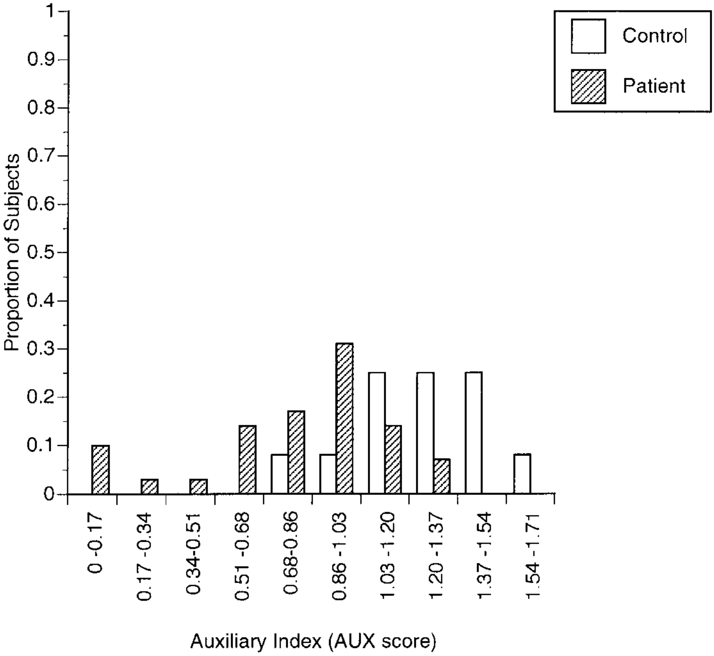

FIG. 6. Frequency distribution of the auxiliary index score ([(Total Auxiliary Score/Number of Matrix Verbs) − 1]) for both groups across the range of scores for that measure.

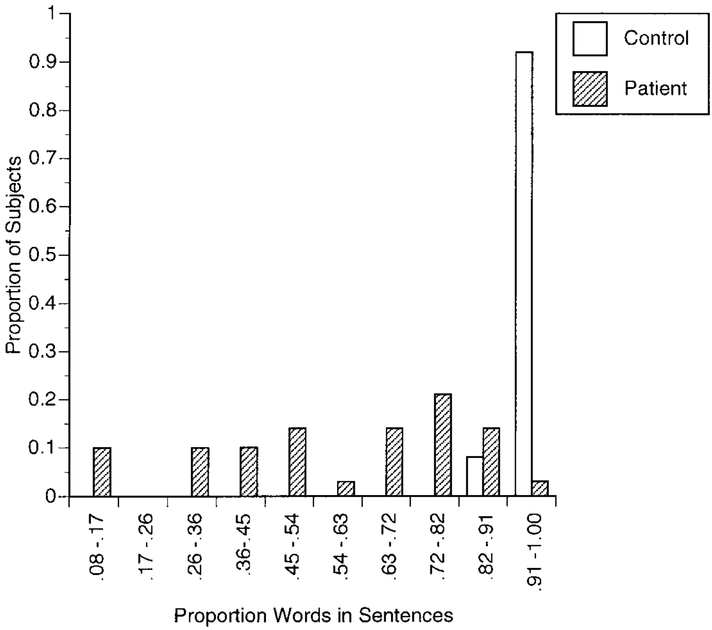

FIG. 7. Frequency distribution of the proportion of narrative words appearing in sentences (Number of Words in Sentences/Number of Narrative Words) for both groups across the range of scores for that measure.

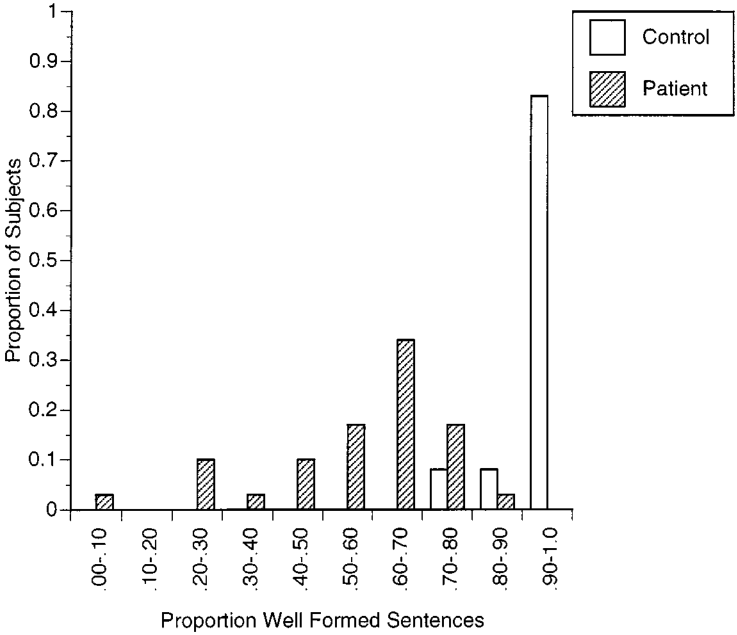

FIG. 8. Frequency distribution of the proportion of well-formed sentences (Number of Well-formed Sentences/Number of Sentences) for both groups across the range of scores for that measure.

contains a verb, and which verb it contains, has strong implications for structural (i.e., structural elaboration of sentences) and aux elaboration. Assuming that this effect is real (a qualification warranted by the relatively low test-retest reliability of the verb measure), it probably reflects the fact that (agrammatic) patients who consistently omit closed class words also tend to underutilize main verbs (Myerson & Goodglass, 1972; Saffran *et al.,* 1980; Bates *et al.,* 1991; Kolk & Heeschen, 1992; Miceli *et al.,* 1984; Zingeser & Berndt, 1990; Hesketh & Bishop, 1996). The fluency variable should correlate with both verb usage and the production of closed class words.

The relationships reflected in Factor 1 are less apparent. While one would expect structural elaboration to pattern with the proportion of words in sentences, and inflections to pattern with aux complexity, it is not clear why these two sets of variables should correlate with one another, although aux complexity could be regarded as another manifestation of structural elaboration. Nor is it the case that a three-factor solution changes these relationships. The only change if three factors are chosen to represent the data is that speech rate (words per minute), alone, comprises the third factor: all other groupings of the variables on the factors remain the same as in the two-factor solution. In any event, it is interesting that free and bound grammatical morphemes

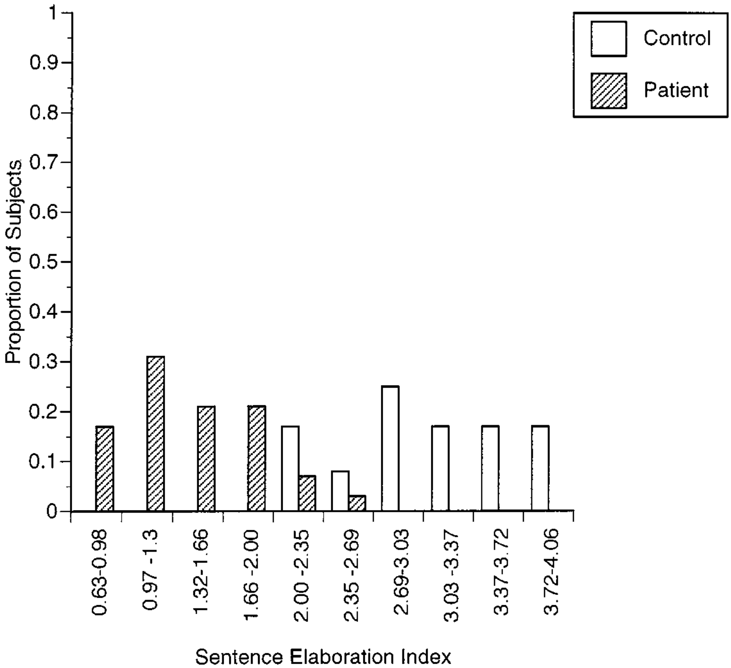

FIG. 9. Frequency distribution of the sentence elaboration index score ([(Number of Words in Subject Noun Phrase/Number of Subject Noun Phrases) − 1] + [(Number of Words in Verb Phrase/Number of Verb Phrases) − 1]) for both groups across the range of scores for that measure.

do not pattern together; nor does the production of free-standing function words pattern with structural complexity.

There is an alternative interpretation of the two-factor structure we obtained. With one exception, the variables in Factor 1 were computed solely on the basis of material that appeared in sentence contexts, whereas those in Factor 2 were computed across the whole narrative sample, including nonsentential utterances. It may be, then, that these factors are reflecting a disposition toward propositional utterances in the case of Factor 1 and ease of word production in the case of Factor 2.

Cluster analysis. To determine whether there were subgroups of patients among the larger group, with different patterns of performance across the two factors, a cluster analysis was performed. For ease of interpretation, individual patient factor scores were submitted to Ward’s (1963) cluster procedure. Briefly, this procedure uses the value of all items in a cluster as a reference point for distances to other clusters, while adjusting for covariances. A four-cluster comparison appeared to offer the most meaningful breakdown. Based upon the group means for the two factor scores, Cluster I ($N = 12$) appears to be the least severe group, overall. This group performs

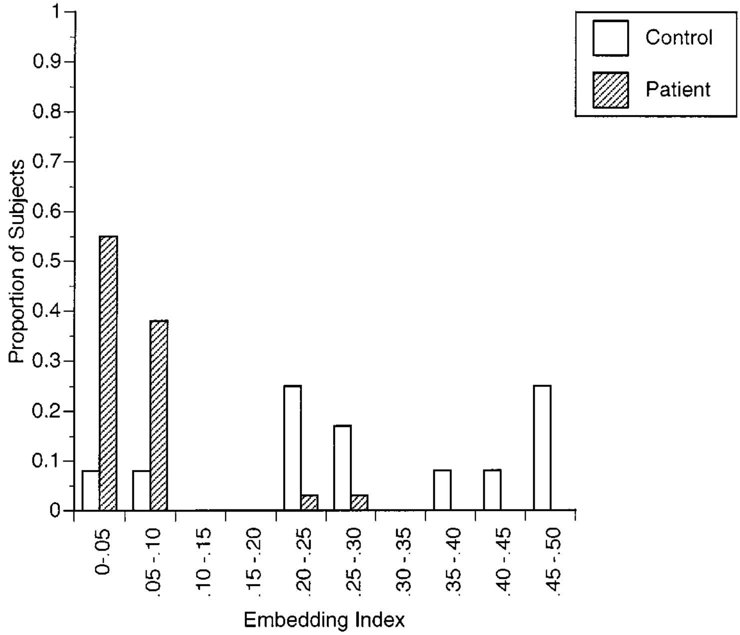

FIG. 10. Frequency distribution of the embedding index score (Number of Embeddings/ Number of Sentences) for both groups across the range of scores for that measure.

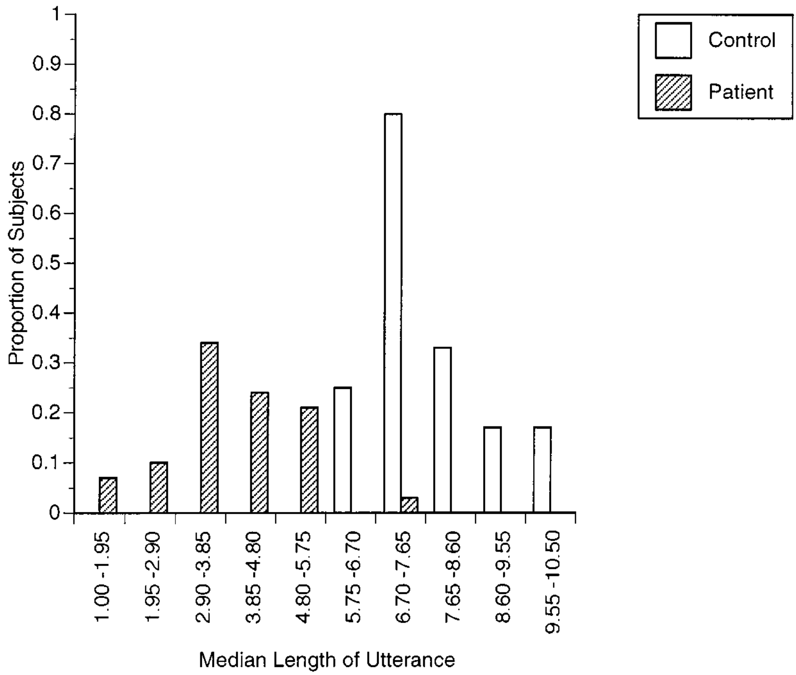

FIG. 11. Frequency distribution of the median length of utterance for both groups across the range of scores for that measure.

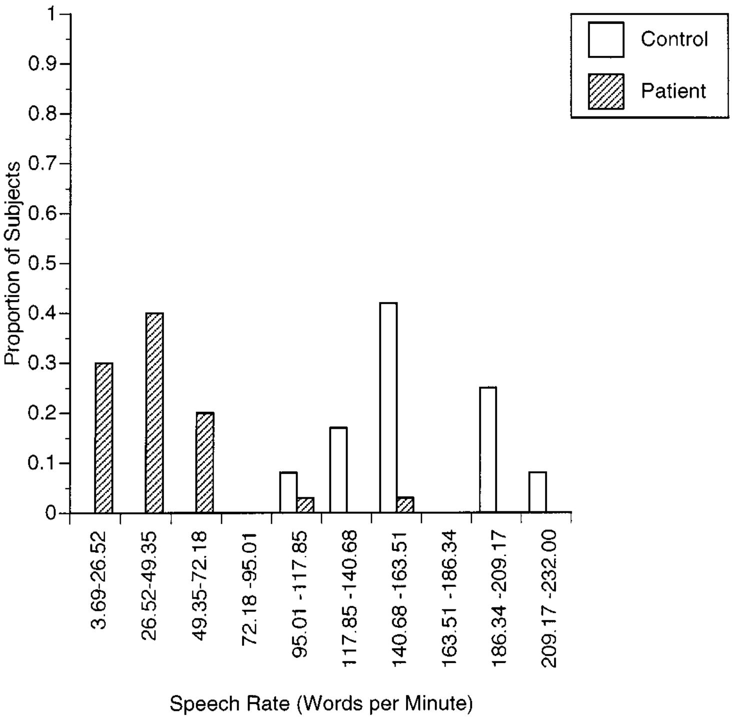

FIG. 12. Frequency distribution of speech rate in words per minute ([(Total Number of Words/Total Time)/60]) for both groups across the range of scores for that measure.

well on both factors. Cluster II ($N = 4$) is the most impaired group overall, performing poorly on both factors. Cluster III ($N = 14$) performs relatively poorly on Factor 1, but better on Factor 2. Cluster IV ($N = 7$) shows the opposite pattern, performing better on Factor 1 than on Factor 2. This analysis indicates that while some patients perform relatively well on both factors (Cluster I), or relatively poorly on both factors (Cluster II), performance on the two factors is also dissociable, as is evident from the different performance patterns of the patients in clusters III and IV. It should be mentioned that the pattern of dissociation just described does not emerge solely in this four-cluster solution. If a three-cluster solution is chosen to represent the data, which is the only other possibility from the results of the cluster analysis, the dissociation in performance between Clusters III and IV remains. Patients in Clusters III and IV do not group together, rather patients from Cluster II join with those in Cluster IV. This finding lends support to the claim that such a dissociation does characterize the data correctly.

Figure 14 illustrates the performance of the Cluster groups and the control subjects across the two factors. For illustrative purposes, the data are presented as standard (z) scores, computed over the entire set of subjects, to

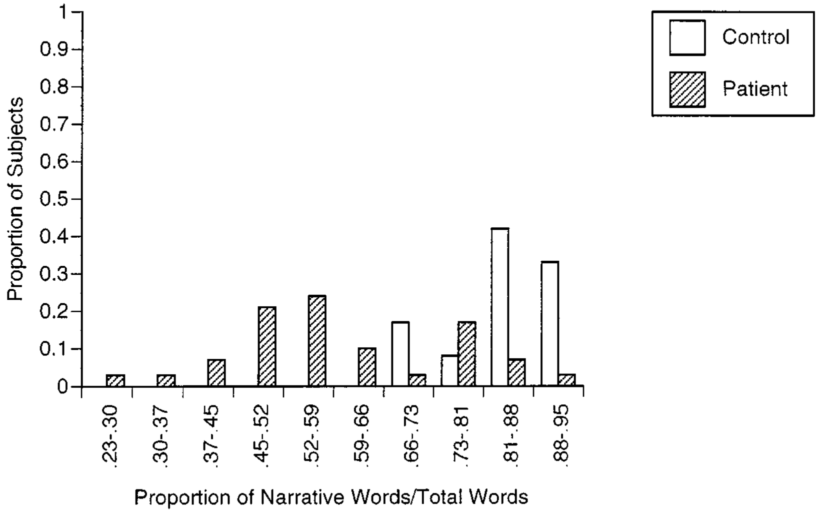

FIG. 13. Frequency distribution of the proportion of narrative to total words produced (struggle) (Number of Narrative Words/Total Number of Words) for both groups across the range of scores for that measure.

equalize scaling among the various measures. For each Cluster, performance (in *z* scores) was averaged across the measures that constituted each of the two factors. Table 6 shows the mean performance of all four of the Cluster groups on all of the individual measures that entered into the factor analysis. Table 6 also includes means for the control subjects on these measures.

Differences between four groups (controls and Clusters I, III, and IV) on

TABLE 5
Principal-Components Factor Analysis

	Varimax (orthogonal) rotation	
QPA Variable	Factor 1	Factor 2
Auxiliary index (Aux.I)	.864	.140
Inflection index (Infl.I)	.807	.155
Proportion words in sentences (W/S)	.796	.478
Sentence elaboration index (SElab)	.696	.236
Proportion verbs (V/N+V)	.358	.804
Determiner index (Det.I)	.354	.761
Proportion closed class words (CC)	.511	.760
Speech rate (WPM)	−.062	.717
Proportion pronouns (P/N+P)	.288	.670
Percentage variance explained	55.600	14.000

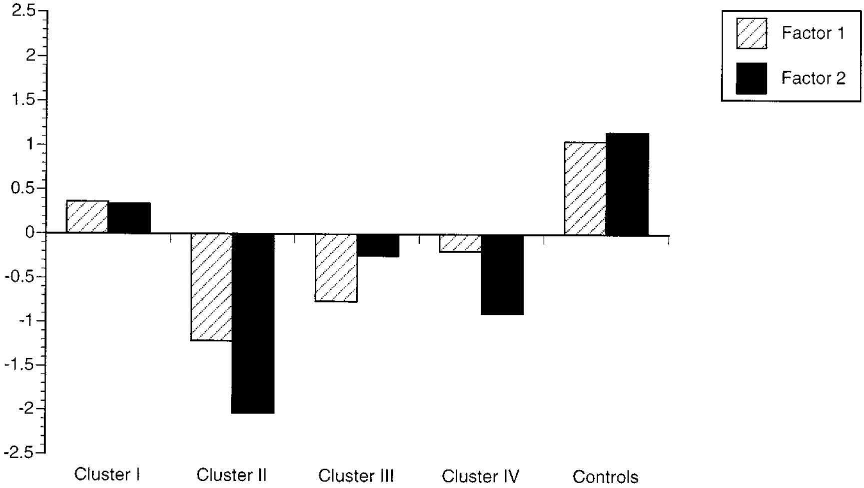

FIG. 14. Mean *z* score profiles for each cluster on the measures contained in Factors 1 and 2.

all the measures included in Factors 1 and 2 were analyzed in separate ANOVAs in order to establish the measures on which the groups differed from each other. Cluster II was not included in this analysis as there were only four subjects in this cluster. For all measures the main effects of group were significant as follows: proportion closed class words ($F(3, 41) = 10.4$, $p < .0001$); determiner index ($F(3, 41) = 15.1$, $p < .0001$); proportion of pronouns to nouns ($F(3, 41) = 7.3$, $p < .001$); proportion of verbs to nouns ($F(3, 41), = 8.4$, $p < .001$); inflection index ($F(3, 41) = 17.1$, $p < .0001$); aux index ($F(3, 41) = 14.4$, $p < .0001$); proportion words in sentences ($F(3, 41) = 32.1$, $p < .0001$); structural elaboration of sentences ($F(3, 41) = 35.4$, $p < .0001$); words per minute ($F(3, 41) = 52.6$, $p < .0001$). Scheffé post hoc tests were used to analyze between-group differences on each measure. Performance of patients in Cluster I was significantly different from control subjects on only two measures: words per minute and the sentence elaboration score. Patients in all cluster groups were significantly slower in their rate of speech (WPM) than control subjects, though they did not differ from each other on this measure. What seems to distinguish patients in Cluster I from control subjects, then, in addition to their slowed speech rate, is their poorer performance on the sentence elaboration measure. This pattern is seen quite clearly upon the inspection of the means in Table 6. Inspection of the composition of Cluster I revealed that all patients but one in this group were considered to be nonfluent nonagrammatic patients by the referring clinicians. This finding is consistent with the results obtained by Saffran *et al.* (1989), who found that the sentence elaboration score was the measure (aside from speech rate) that most differentiated the nonagrammatic patients from the control subjects.

TABLE 6
Group Means for Individual Measures in Factors 1 and 2 for Aphasic Patients in Each Cluster and Control Subjects

Factor 1	Cluster I ($N = 12$)	Cluster II ($N = 4$)	Cluster III ($N = 14$)	Cluster IV ($N = 7$)	Controls ($N = 12$)
Auxiliary index (Aux.I)					
Mean	1.070	.583	.586	.919	1.260
SD	(.282)	(.359)	(.293)	(.212)	(.26)
Inflection index (Infl.I)					
Mean	.783	.340	.401	.666	.920
SD	(.126)	(.320)	(.278)	(.135)	(.15)
Proportion words in sentences (W/S)					
Mean	.878	.228	.512	.584	.980
SD	(.109)	(.135)	(.183)	(.173)	(.05)
Sentence elaboration index (SElab)					
Mean	1.786	.755	1.244	1.464	3.060
SD	(.458)	(.231)	(.282)	(.476)	(.64)
Factor 2					
Proportion verbs (V/N+V)					
Mean	.454	.160	.391	.323	.480
SD	(.054)	(.091)	(.096)	(.056)	(.06)
Determiner index (Det.I)					
Mean	.888	.140	.644	.524	.990
SD	(.085)	(.201)	(.255)	(.236)	(.02)
Proportion closed class words (CC)					
Mean	.501	.217	.409	.396	.540
SD	(.048)	(.082)	(.112)	(.044)	(.04)
Speech rate (WPM)					
Mean	42.901	18.425	42.923	23.939	160.820
SD	(15.562)	(2.01)	(14.094)	(6.989)	(37.00)
Proportion pronouns (P/N+P)					
Mean	.343	.057	.312	.091	.410
SD	(.117)	(.045)	(.220)	(.041)	(.09)

The composition of the patients in Clusters III and IV is quite different from that of Cluster I: in each of the two groups all patients but one were clinically diagnosed as agrammatic (and the four patients in Cluster II, not included in this analysis, were all considered to be severely agrammatic). For each of these two groups, the pattern of performance in terms of the differences between the patient group and control subjects was dramatically different than for Cluster I. Cluster III, the group that performed better overall on Factor 2 than Factor 1, performed significantly more poorly on all measures compared to control subjects with the exception of the proportion of pronouns measure (a Factor 2 measure). Cluster IV, the group that performed better on Factor 1 than on Factor 2, performed significantly more poorly on all measures compared to control subjects with the exception of the auxiliary index measure (a Factor 1 measure). The different pattern of performance between Clusters III and IV on the variables that comprise Factor 1 and Factor 2 can be seen upon inspection of the means in Table 6.

Scheffé comparisons between the patient groups indicated that Cluster III's performance was poorer than Cluster I's on all Factor 1 variables and on two Factor 2 variables (proportion closed class words and determiner index), but not on the three remaining variables that Factor 2 loaded on: inflection index, words per minute, and verbs as a proportion of nouns and verbs. This is as would be expected as Cluster III is the group that performed relatively better on Factor 2 than on Factor 1. For Cluster IV, the group that showed the opposite pattern of performance, performance was poorer than that of Cluster I patients on all Factor 2 measures except words per minute, whereas the two groups did not differ on any Factor 1 variables except the proportion of words in sentences measure. Clusters III and IV differed from each other only on the pronoun measure (Cluster III performing better than Cluster IV) and the inflection index measure (Cluster IV performing better than Cluster III).

The above comparisons between the performance of the patients in the different clusters on the QPA measures included in the principle components analysis highlight two characteristics of the data. The first is the heterogeneity of impairment that emerges in patients characterized as nonfluent Broca's aphasics with respect to the morphological and structural indices that are measured in the QPA. The data suggest that there are patients with different patterns of deficits. The second characteristic is that of continuity between all subject groups on the production indices, both morphological and structural, in this study. While the patients and control subjects were statistically distinguishable on all measures, as seen in Table 4, Figs. 1 to 13 clearly illustrated that there was some degree of overlap on all measures between the two groups. Comparisons between the cluster groups extend this finding. One patient group, the highest performing, differed from controls on only one measure, whereas the two groups that showed dissociation of performance on the two Factors tended to differ from the control subjects only on the measures comprising one of the Factors. Similarly, these two groups tended to differ from the highest performing group primarily on the measures in the Factor that they performed most poorly on and not the others. The group that performed poorly overall was not included in these comparisons, but it is reasonable to speculate that this group differed from the controls, at least, on all measures.

GENERAL DISCUSSION

For the most part, the results of this study affirm our confidence in the reliability of the QPA measures. There are a few exceptions. Number of verbs and embeddings proved to have rather low reliability for the control subjects. The assessment of embeddings, particularly in the complex sentences produced by many of the controls, can pose a challenge for scorers with limited syntactic sophistication. To remedy this, we have prepared a manual that contains many more examples of embedded structures (Berndt *et al.,* 2000). The verb count may reflect a similar problem. The verb measure (i.e., verbs/noun + verbs) and the well-formedness measures had relatively low test-retest reliability for the patients. The verb measure includes words that occurred outside sentences; assigning words to syntactic categories is therefore likely to be a problem in patient samples, where words often occur outside sentence contexts. As there is significant overlap between nouns and verbs in English (e.g., dance, dress, ride), we dealt with this ambiguity by asking scorers to consult dictionary listings for these words, using the first listing (noun or verb) to classify the item. As noted, another source of difficulty may be the repeated use of a word for emphasis (e.g., dance, dance, dance), which occurred relatively frequently in the patient samples; in these instances, we did not remove repetitions as we did for multiple attempts at a word in selecting the narrative sample. In the case of the well-formedness measure, the low reliability is very likely due to the limited number of ''sentences'' produced by the aphasic subjects.

With respect to relationships among measures, the principal components analysis revealed a degree of independence between the production of free-standing function words and the structural measures, which patterned with inflections and aux complexity. Furthermore, the cluster analysis yielded two sets of patients (Clusters III and IV) who showed different performance patterns across these measures. The patients in Cluster III performed better on free-standing function words than on the structural and bound morpheme measures, while those in Cluster IV showed the opposite pattern. These results substantiate the indications in our earlier data from individual patients (Saffran *et al.,* 1989) that the production of free and bound grammatical morphemes is dissociable (also see Miceli *et al.,* 1989). They also support findings from single case analyses that suggest that syntactic complexity can dissociate from function word production (e.g., Berndt, 1987; Nespoulous *et al.,* 1988). Taken together, these data provide support for models of sentence production in which sentence structure, in the form of a frame marked with slots for particular classes of words (e.g., noun, verb) subsequently to be inserted, does not contain full specifications for grammatical morphemes. It appears, rather, these must be retrieved in a separate operation, which may be impaired independently (see LaPointe & Dell, 1989, for one suggestion along these lines).

As in the Saffran *et al.* (1989) study, we found that nonfluent patients differed from controls with respect to the structural elaboration of the sentences they produced, irrespective of their classification as agrammatic or nonagrammatic. The explanation of this finding remains obscure. It is possible that it reflects a syntactic limitation that is present in all nonfluent patients; alternatively, the structural deficit could be a manifestation of a slow rate of word retrieval or possibly a capacity or working memory decrement associated with frontal lesions. Using the QPA analyses, Bird and Franklin

(1996) found decreased syntactic elaboration in two fluent patients; in one patient whose recovery pattern was documented, syntactic elaboration increased along with improvement in word retrieval. It is reasonable to assume that there is some relationship between the ability to retrieve words—and to be able to maintain at least a few of them concurrently—and the ability to produce complex sentences (for a recent formulation of the relationship between sentence production and lexical-semantic retention deficits, see Martin *et al.,* 1998).

To summarize the major results of this study: (1) We have demonstrated the reliability of the QPA and most of the measures that it yields. (2) We have demonstrated that nonfluent patients, whether they satisfy clinical criteria for agrammatism or not, differ from controls in the structural elaboration of the sentences they produce, as well as on the speech rate variable which entered into the criteria for selection. (3) We have provided additional evidence that agrammatic patients can differ in their production of bound and free grammatical morphemes, and that the former appears more closely associated with the elaboration of sentence structure.

ACKNOWLEDGMENTS

The preparation of this manuscript was supported by Heart and Stroke Foundation of Canada Grant NA3175 to Elizabeth Rochon, and NIH Grants DC 000191 to Eleanor Saffran and DC 00262 to Rita Berndt. We thank Dr. Randi Martin for contributing two patient narratives. Thanks also to Gwen Dorfman, Dr. Nadine Martin, Jessica Myers, Paula Sobel, and Sarah Wayland for help in scoring transcripts. Thanks to Malcolm Binns, Rotman Research Institute, Baycrest Centre for Geriatric Care, for statistical consultations.

REFERENCES

Bates, E., Chen, S., Tzeng, O., Li, P., & Opie, M. 1991. The noun–verb problem in Chinese. *Brain and Language,* **41,** 203–233.

Berndt, R. S. 1987. Symptom co-occurrence and dissociation in the interpretation of agrammatism. In M. Clotheart, R. Job, and G. Sartori (Eds.), *The cognitive neuropsychology of language* (pp. 221–233). London: Erlbaum.

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E., & Schwartz, M. (2000). *Quantitative production analysis: A training manual for the analysis of aphasic sentence production.* Hove, U.K.: Psychology Press.

Berndt, R. S., Mitchum, C. C., Haendeges, A. N., & Sandson, J. 1997. Verb retrieval in aphasia: 1. Characterizing single word impairments. *Brain and Language,* **56,** 68–106.

Berndt, R. S., Haendeges, A. N., Mitchum, C. C., & Sandson, J. 1997. Verb retrieval in aphasia: 2. Characterizing single word impairments. *Brain and Language,* **56,** 107–137.

Bird, H., & Franklin, S. 1996. Cinderella revisited: A comparison of fluent and non-fluent aphasic speech. *Journal of Neurolinguistics,* **9(3),** 187–206.

Bock, J. K. 1986. Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **12,** 575–586.

Byng, S. 1988. Sentence processing deficits: Theory and therapy. *Cognitive Neuropsychology,* **5,** 629–676.

Byng, S., Nickles, L., & Black, M. 1994. Replicating therapy for mapping deficits in agrammatism: remapping the deficit? *Aphasiology,* **8(4),** 315–341.

Edwards, S. 1995. Profiling fluent aphasic spontaneous speech: A comparison of two methodologies. *European Journal of Disorders of Communication,* **30,** 333–345.

Goodglass, H., & Kaplan, E. 1983. *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia, PA: Lea & Febiger.

Hesketh, A., & Bishop, D. V. M. 1996. Agrammatism and adaptation theory. *Aphasiology,* **10(1),** 49–80.

Kolk, H. H. J., & Heeschen, C. 1992. Agrammatism, paragrammatism, and the management of language. *Language and Cognitive Processes,* **7,** 89–129.

LaPointe, S. G., & Dell, G. S. 1989. A synthesis of some recent work in sentence production. In G. N. Carlson and M. K. Tanenhaus (Eds.) *Linguistic structure in language processing* (pp. 107–156). Dordrecht: Kluwer.

Martin, R. C., Katz, M., & Freedman, M. 1998. Lexical-semantic retention and language production. *Brain and Language,* **65(1),** 99–101.

Martin, R. C., Wetzel, F., Blossom-Stach, C., & Feher, E. 1989. Syntactic loss versus processing deficit: An assessment of two theories of agrammatism and syntactic comprehension deficits. *Cognition,* **32(2),** 157–191.

Miceli, G., Silveri, M. C., Romani, C., & Caramazza, A. 1989. Variation in the pattern of omissions and substitutions of grammatical morphemes in the spontaneous speech of so-called agrammatic patients. *Brain and Language,* **36,** 447–492.

Miceli, G., Mazzucchi, A., Menn, L., & Goodglass, H. 1983. Contrasting cases of Italian agrammatic aphasia without comprehension disorder. *Brain and Language,* **19,** 65–97.

Miceli, G., Silveri, M. C., Villa, G., & Caramazza, A. 1984. On the basis for the agrammatic's difficulty in producing main verbs. *Cortex,* **20,** 207–220.

Myerson, R., & Goodglass, H. 1972. Transformation grammars of three agrammatic patients. *Language and Speech,* **15,** 40–50.

Nespoulous, J.-L., Dordain, M., Perron, C., Ska, B., Bub, D., Caplan, D., Mehler, J., & Lecours, A. R. 1988. Agrammatism in sentence production without comprehension deficits: Reduced availability of syntactic structures and/or of grammatical morphemes? A case study. *Brain and Language,* **33,** 273–295.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. 1989. The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language,* **37,** 440–479.

Saffran, E. M., Schwartz, M. F., & Marin, O. S. M. 1980. Evidence from aphasia: Isolating the components of a production model. In B. Butterworth (Eds.), *Language production* (pp. 221–241). London: Academic Press.

Schwartz, M. F., Saffran, E. M., Fink, R. B., Myers, J. L., & Martin, N. 1994. Mapping therapy: A treatment programme for agrammatism. *Aphasiology,* **8(1),** 19–54.

Snedecor, G. W., & Cochran, W. G. 1989. *Statistical methods* (8th ed.). Ames, IA: Iowa State University Press.

Tabachnick, B. G., & Fidell, L. S. 1989. *Using multivariate statistics* (2nd ed.). New York, NY: HarperCollinsPublishers, Inc.

Ward, J. 1963. Hierarchical grouping to optimize on function. *American Statistical Association Journal,* **58,** 236–244.

Zingeser, L., & Berndt, R. S. 1990. Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language,* **39,** 14–32.