

# A psychometric analysis of functional category production in English agrammatic narratives

Lisa H. Milman<sup>a,\*</sup>, Michael Walsh Dickey<sup>b,c</sup>, Cynthia K. Thompson<sup>b,c,d,e</sup>

<sup>a</sup> Department of Speech and Hearing Science, Ohio State University, 1070 Carmack Road, Columbus, OH 43210, USA

<sup>b</sup> Aphasia and Neurolinguistics Research Laboratory, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA

<sup>c</sup> Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA

<sup>d</sup> Department of Neurology, Northwestern University, 710 North Lake Shore Drive, Chicago, IL 60611, USA

<sup>e</sup> Cognitive Neurology and Alzheimer's Disease Center, Northwestern University, 320 E. Superior, Chicago, IL 60611, USA

Accepted 12 December 2007

Available online 5 February 2008

## Abstract

Hierarchical models of agrammatism propose that sentence production deficits can be accounted for in terms of clausal syntactic structure (Friedmann & Grodzinsky, 1997; Hagiwara, 1995). Such theories predict that morpho-syntactic elements associated with higher nodes in the syntactic tree (complementizers and verb inflections) will be more impaired than elements associated with lower structural positions (negation markers and aspectual verb forms). While this hypothesis has been supported by the results of several studies (Friedmann & Grodzinsky, 1997; Hagiwara, 1995), it has also been challenged on several grounds (Burchert, Swoboda-Moll, & De Bleser, 2005a; Lee, 2003; Lee, Milman, & Thompson, 2005). In this paper the question of hierarchical structure was re-examined within the framework of Item Response Theory (IRT, Rasch, 1980). IRT is a probabilistic model widely used in the field of psychometrics to model behavioral constructs as numeric variables. In this study we examined production of functional categories (complementizers, verb inflections, negation markers, and aspectual verb forms) in narrative samples elicited from 18 individuals diagnosed with nonfluent aphasia and 18 matched controls. Data from the aphasic participants were entered into an IRT analysis to test (1) whether production of clausal functional categories can be represented as a variable on a numeric scale; and (2) whether production patterns were consistent with hierarchical syntactic structure. Pearson *r* correlation coefficients were also computed to determine whether there was a relation between functional category production and other indices of language performance. Results indicate that functional category production can be modeled as a numeric variable using IRT. Furthermore, although variability was observed across individuals, consistent patterns were evident when the data were interpreted within a probabilistic framework. Although functional category production was moderately correlated with a second measure of clausal structure (clause length), it was not correlated with more distant language constructs (noun/verb ratio and WAB A.Q.). These results suggest that functional category production is related to some, but not all, measures of agrammatic language performance.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** Aphasia; Agrammatism; Item Response Theory; Psychometric; Quantitative; Functional categories; Morphology; Syntax

## 1. Introduction

Over the last decade much debate has focused on the extent to which the varied manifestations of agrammatism observed across languages can be explained in reference to clausal syntactic structure. Within the framework of gener-

ative syntax, a clause consists of lexical and functional phrases arranged in hierarchical fashion (see Fig. 1). Lexical phrases, such as the Verb Phrase (VP), located in the lower portion of the tree structure, are associated with basic lexical categories including nouns, verbs, adjectives, and adverbs. In contrast, functional phrases, such as

\* Corresponding author. Fax: +1 614 292 7504.

E-mail address: [milman.3@osu.edu](mailto:milman.3@osu.edu) (L.H. Milman).

Aspect Phrase (ASPP), Negation Phrase (NegP), Inflectional Phrase (IP), and Complementizer Phrase (CP), positioned progressively higher in the tree, are associated with specific grammatical or functional categories including aspectual verb forms (*\_ing, \_en*), negation markers (*not, never*), finite verb inflections (*tense, agreement*), and complementizers (*if, whether, that*).

Several researchers (Friedmann & Grodzinsky, 1997; Hagiwara, 1995; Ouhalla, 1993) have suggested that the hierarchical structure of syntactic trees may underlie the varied patterns of impairment observed in agrammatism. According to such hierarchical theories, functional categories associated with higher nodes in the tree structure are selectively impaired in agrammatism. While impairment may occur at various structural levels, an impairment at any given level necessarily entails impairment at all higher levels. Thus if the IP node is affected, it follows that the CP node must also be affected since it is positioned higher in the clausal structure.

Empirical support for this hypothesis has come from multiple studies of agrammatism conducted in several languages, using a variety of production tasks. For instance, since this hierarchical arrangement situates nouns and nonfinite verbs in a lower (and hence relatively protected) structural position, it captures two widely accepted generalizations regarding agrammatism: the preservation of open class (lexical) relative to closed class (grammatical/functional) morphology (Caramazza & Berndt, 1985;

Goodglass & Menn, 1985; Saffran, Berndt, & Schwartz, 1989); and the relative preservation of nonfinite relative to finite verb forms (Goldstein, 1948; Menn & Obler, 1990; Myerson & Goodglass, 1972). Moreover, several recent studies have found production of elements associated with CP to be more impaired than production of elements associated with lower functional projections. This selective impairment of CP has been observed in Japanese (Hagiwara, 1995); Hebrew, Palestinian Arabic, and English (Friedmann, 2001,2002); German (Burchert, Swoboda-Moll, & De Bleser, 2005b) and Greek (Stavrakaki & Kouvava, 2003). Consistent with the tree structure proposed by Pollock (1989), in which IP is ‘split’ further into separate tense and agreement nodes, the Tree Pruning Hypothesis (TPH) has also proposed that verb morphology associated with tense should be more impaired than morphology associated with agreement. Supporting data have come from Hebrew and Palestinian Arabic (Friedmann, 2001; Friedmann & Grodzinsky, 1997), English (Goodglass & Berko, 1960), Spanish (Benedet et al., 1998; Milman, 1997) and German (Wenzlaff & Clahsen, 2005).

Three types of evidence, however, seem to contradict such hierarchical accounts. First, cases have been reported in Korean (Lee, 2003), German (Burchert et al., 2005a), and English (Lee et al., 2005) showing impairment patterns that are opposite to those predicted by a hierarchical model. For example, Lee et al. (2005) describe four English

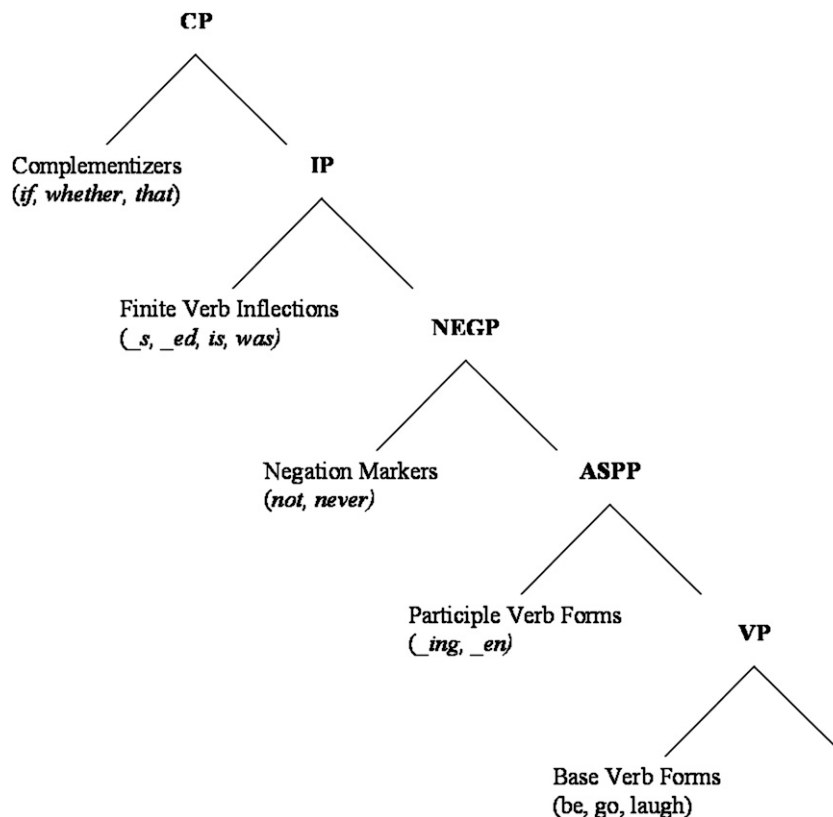


Fig. 1. Syntactic tree structure (modeled after Ouhalla, 1991).

agrammatic patients with spared CP morphology but impaired production of IP related morphology.

A second issue concerns the nature (or extent) of the impairment. In many cases performance is characterized by partial knowledge rather than a complete loss of ability. This is evident in the variable range of scores observed across tasks and subjects as well as in the occurrence of a variety of error types, including both omission and substitution errors. The fact that agrammatic patients demonstrate partial knowledge of a syntactic structure seems incompatible with the notion that the representation is simply absent or ‘pruned’ from the grammar. Rather it suggests that the impairment is more likely due to faulty processing or access to the relevant representation (see for example discussions in Arabatzi & Edwards, 2002; Stavrakaki & Kouvava, 2003).

Lastly, there is a substantial body of evidence indicating that other factors, besides hierarchical clausal structure, are relevant to understanding agrammatic production. Examples include: dissociations between morphological and syntactic processing (Bastiaanse & Thompson, 2003; Miceli, Mazzucchi, Menn, & Goodglass, 1983), between free and bound grammatical morphology (Rochon, Saffran, Berndt, & Schwartz, 2000; Thompson, Fix, & Gitelman, 2002), and variability (or underspecification) within individual phrases (Burchert et al., 2005a; Hagiwara, 1995). Notably, the noun phrase (or determiner phrase) has been described as having its own hierarchical structure that is *independent* of clausal structure (see discussions in Abney, 1987; Cinque, 2002; Haegeman, 1994).

The empirical record thus leaves us with the familiar problem of having a theoretical model that accounts for some of the data but is unable to account for all of it. Traditionally, in the agrammatism literature, this type of impasse has been dealt with in two ways: either by presenting an account that explains more of the data (see for example Arabatzi & Edwards, 2002; Burchert et al., 2005a) or by proposing that the performance variability observed across patients precludes any general description (Badecker & Caramazza, 1985; McCloskey, 1993; Miceli, Silveri, Romani, & Caramazza, 1989). In this paper we take a different approach to the issue of performance variability and recast the question within a psychometric framework. Specifically, we use Item Response Theory (IRT, Rasch, 1980; Wright & Stone, 1979) to ask whether there is any psychometric basis for treating clausal structure as a single variable contributing to agrammatic production performance. While psychometric approaches have been applied to a variety of agrammatic comprehension (Bates, McDonald, MacWhinney, & Appelbaum, 1991; Drai & Grodzinsky, 2006) and production data (Rochon et al., 2000; Saffran et al., 1989; Thompson et al., 1995), we extend this approach and use IRT to investigate whether clausal functional categories can be grouped together hierarchically as points defining a single variable.

Item Response Theory (IRT) is widely used in the field of psychometrics to mathematically specify variables related to

human cognitive performance. IRT has been used to develop scholastic aptitude tests such as the Graduate Record Exam (GRE, Briel, O’Neill, & Scheuneman, 1993) and the Medical College Admission Test (MCAT, Koenig, 1998). It has also been used to develop and evaluate measurement scales for a variety of neuropsychological tools including the Wechsler Adult Intelligence Scales (WAIS, Wechsler, 1997a), the Wechsler Memory Scales (WMS, Wechsler, 1997b), the Revised Token Test (Hula, Doyle, McNeil, & Mikolic, 2006), and a variety of functional communication measures (Doyle et al., 2004; Granger, Hamilton, Linacre, Heine-mann, & Wright, 1993; Milman & Holland, 2004).

IRT is a useful model for addressing the questions posed in this paper for several reasons. Although we gain some ground simply by counting the number of omissions and substitutions of grammatical morphemes, the use of an inferential statistical model, such as IRT, makes it possible to go beyond describing the data at hand. It allows us to generate and test hypotheses about abstract psychological constructs and further to predict the performance of larger populations. One important feature of the IRT model in this regard is its emphasis on precise definition of the variable that is being measured. Specifically, IRT aims to define variables in terms of discrete behaviors (or items) that are ordered hierarchically in terms of difficulty. For instance, basic mathematical ability might be defined in terms of performance on the following three items:

- i.  $2 + 4$
- ii.  $27/9$
- iii.  $2^4/8$

This emphasis on items (or behaviors) to define cognitive variables also provides the foundation for more objective measurement. Rather than evaluating ability based on the performance of a small group of individuals (how some individuals perform relative to others), ability can be evaluated in terms of a set of theoretically motivated skills (e.g., addition, division, exponential functions, etc.). In essence, IRT creates a yardstick to objectively measure particular cognitive abilities.

The mathematical properties associated with IRT provide a further set of advantages. Not only are variables defined conceptually as a set of ordinally sequenced skills, but also *numerically* as a series of points on an equal interval scale. Critically, each item on the IRT scale is associated with a specific numeric value. Furthermore, since items and individual ability are measured on the same scale, the model can generate precise and falsifiable predictions about how individuals of a particular ability will perform on any given item. Moreover, the numeric properties of the IRT scale make it possible to statistically identify misfitting items or individuals that are incongruent with the scale, and hence our model of the construct. A further advantage of IRT is that it is a probabilistic rather than deterministic model. In fact, it was developed

explicitly to model the response variability that is so typical of human cognitive performance.

The standard representation of the IRT scale is shown in Fig. 2. Note that the equal interval IRT scale is situated in the center of the figure. By convention, the midpoint of the scale is set at a value of “0”. The items used to define the construct are listed to the right of the scale and are ordered in terms of their difficulty ( $\delta$ ) from easy (bottom) to hard (top). The scores of individuals are plotted to the left of the scale and are ordered in terms of ability ( $\beta$ ) from low (bottom) to high (top).

A critical assumption of the IRT model is that the difference between the ability of an individual and the difficulty of an item ( $\beta - \delta$ ) can be used to predict response probabilities. Stated differently, when individuals and items fit a scale, individuals are likely to succeed on items that are below their ability level and fail on items that are above their ability level. For instance, in Fig. 2 Person<sub>w</sub> is likely to succeed on items *k*, *l* and *m* (below ability level) but fail item *i* (above ability level).

Hierarchical accounts of agrammatism predict that functional categories can be represented on such a scale with higher nodes (e.g., COMP and INFL) depicted as more difficult items, and lower nodes (e.g., NEG and ASP) as easier items (see Fig. 3).

If this representation is accurate, we would expect to see performance patterns similar to those shown in persons 1–3 depicted in Table 1. For example, an individual with a low level of ability (see person 1) is likely to suc-

ceed on the easiest item (in this case aspectual verb forms), but is more likely to miss items as they become progressively more difficult (negation markers, verb inflections, complementizers). Similarly, we would expect person 2 to produce items that are matched to their ability level (aspectual verb forms and negation markers), but not items that are above this level (verb inflections and complementizers).

Given the variability that is characteristic of human performance, we might also expect to see production patterns as shown in persons 4, 5, and 6. In this case most of the data points match the predicted outcomes, although the performance pattern of person 5 departs somewhat from these expectations. The performance patterns of persons 7, 8, and 9, however, are clearly inconsistent with the predicted patterns.

Of course it would also be possible for only a subset of items to pattern together. For example, in persons 10, 11, and 12, a hierarchical pattern is observed for complementizers, inflections, and negation markers. However, a hierarchical pattern is not observed for aspectual verb forms. Specifically, person 12 (highest level of ability in this subgroup) missed this item; whereas person 10 (lowest level of ability in this subgroup) was successful on this item. If clausal functional categories do in fact constitute a natural set of grammatical morphemes, we would expect morphemes that are *not* associated with clausal structure (such as plural “\_s” marking on noun phrases) to show exactly this misfitting pattern.

In this paper we adopt a psychometric approach to investigate clausal functional category production in agrammatic narratives. Our primary goal is to evaluate whether clausal functional categories can be grouped

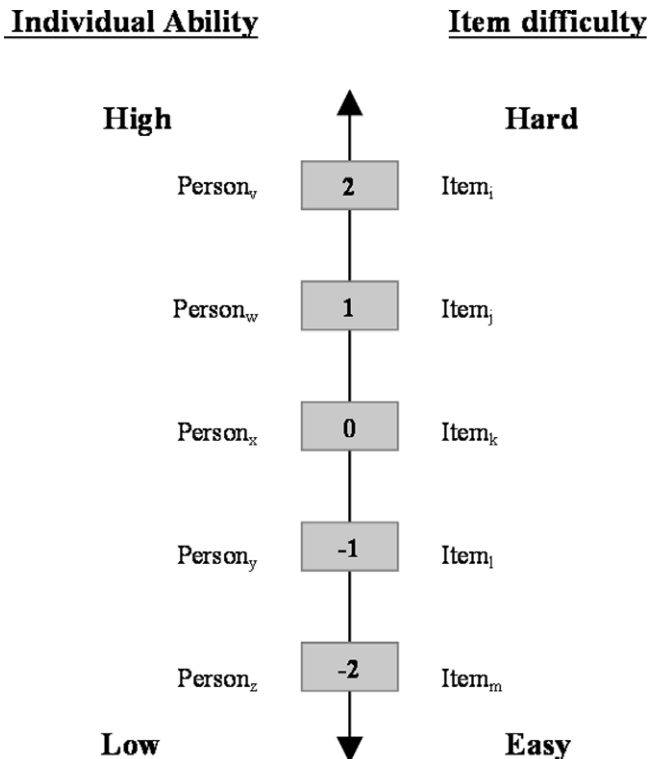


Fig. 2. IRT measurement scale depicting individual ability and item difficulty.

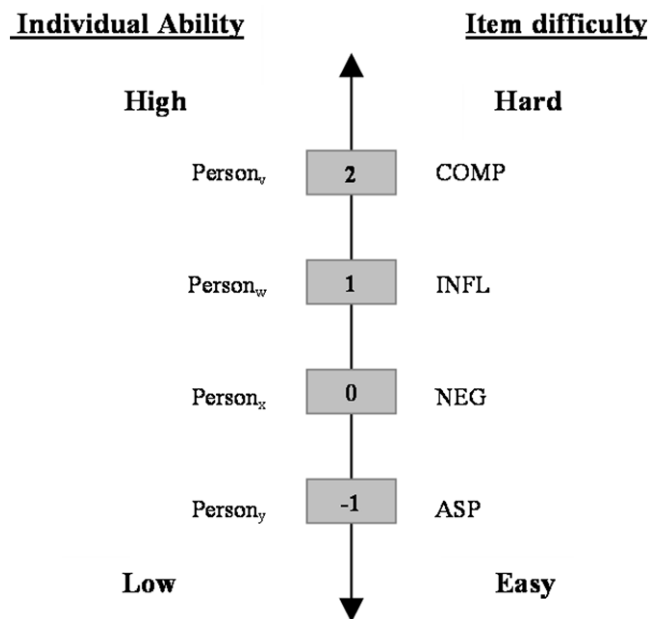


Fig. 3. Hypothetical measurement scale for hierarchical models of agrammatism.

Table 1  
Possible production patterns for functional categories

	COMP	INFL	NEG	ASP	Ability score ( $\beta$ ) (sum of scores for each person)
Person 1	0	0	0	1	1
Person 2	0	0	1	1	2
Person 3	0	1	1	1	3
Person 4	0	0	0	1	1
Person 5	0	0	1	0	1
Person 6	0	1	1	1	3
Person 7	1	0	0	0	1
Person 8	1	0	1	0	2
Person 9	1	0	0	1	3
Person 10	0	0	1	1	2
Person 11	0	1	1	0	2
Person 12	1	1	1	0	3
Difficulty score ( $\delta$ ) (sum of scores for each item)	4	4	8	7	

Note. 0, impaired ability; 1, preserved ability.

together as a series of items defining a single variable. A second purpose is to determine whether the order of impairment observed across functional categories is consistent with predictions made by hierarchical accounts (CP > IP > NegP > AspP). A final objective is to examine the relation between clausal functional category production and other aspects of language processing.

If hierarchical accounts such as the TPH are correct, we would expect to see a relatively limited and fixed set of performance patterns, as in persons 1–3 (see Table 1) in which an impairment at one level necessarily entails an impairment at all higher levels. In contrast, if the effect of clausal syntactic structure is variable across all patients, we would expect to see random variation as in persons 7–9. Another possibility, proposed here, is that performance on clausal structures does follow a consistent pattern; however, this pattern is not absolute, and is more accurately expressed in probabilistic terms (as in persons 4, 5, and 6). IRT provides a well-established psychometric model for evaluating these hypotheses.

We begin by surveying production of clausal (complementizers, tense inflection, agreement inflection, negation markers, aspectual verb forms) and nonclausal (plural “\_s” marking on nouns) functional categories in the narrative samples (Cinderella stories) of 18 neurologically healthy controls and 18 individuals diagnosed with agrammatic aphasia. Normative data from the control group are used to derive standardized scores to assess functional category production in the aphasic participants. These scores are entered into an IRT analysis to determine whether clausal functional categories pattern as a coherent set, and whether there is any psychometric basis for treating clausal hierarchical structure as a variable characterizing agrammatic sentence production.

Lastly, IRT scores are correlated with other language measures to evaluate the relation between functional category production and more general language performance.

## 2. Methods

### 2.1. Participants

Eighteen neurologically healthy controls and 18 individuals diagnosed with nonfluent aphasia participated in this research. All participants were native English speakers and had no current/pre-morbid history of language, learning, or neurological disorder. The diagnosis of nonfluent aphasia was based on documented evidence of a single left hemisphere stroke in neurological/radiological reports, performance patterns on the Western Aphasia Battery (WAB, Kertesz, 1982), and an overall clinical impression of effortful agrammatic speech with relatively well preserved language comprehension. Aphasic participants (11 male and 7 female) were at least 3 months post onset when the language samples were collected and demonstrated mild to moderate impairments of general language ability on the WAB (mean Aphasia Quotient, AQ, was 70.9 and ranged from 52.3 to 85.2). Aphasic participants ranged in age from 29 to 78 years (mean = 54 years) and had 12–20 years of education (mean = 15 years). All but one of the patients was right handed. WAB AQ, age, education, gender, and handedness for each patient are shown in Table 2. Demographic information was not available for 2 of the 18 neurologically healthy controls. For the remaining 16 subjects (8 male and 8 female) age ranged from 28 to 86 years (mean = 56 years), and years of education ranged from 14 to 24 (mean = 18 years).

### 2.2. Procedures

#### 2.2.1. Narrative data collection and coding

Narrative language samples were drawn from a pre-existing data bank collected between 1993 and 2005 at the Northwestern University Aphasia and Neurolinguistics Research Laboratory. Participants were shown a picture book depicting the story of Cinderella. After viewing the pictures, the book was removed and subjects were asked to tell the story. Narrative samples were transcribed, segmented into utterances, and coded following procedures developed by Thompson and colleagues (Thompson et al., 1995). Utterances were further segmented into individual clauses containing minimal predicate structure (e.g., ‘Animals help out too’, ‘go to the ball’, ‘dancing’, ‘Cinderella and prince happy’). Each clause was then evaluated for the inclusion or omission of the following clausal functional categories: complementizers, verb inflections (past tense, agreement, infinitive ‘to’, modal verbs), negation markers, and aspectual verb form. In addition, morphemes associated with plural marking were tallied. This was done to assess whether all gram-

Table 2  
Demographic information for aphasic participants

Participant	WAB A.Q. Mean = 70.9 SD = 9.8	Age in years Mean = 54 SD = 13.8	Education in years Mean = 14.9 SD = 1.9	Gender M:F = 11:7	Handedness R:L = 14:1
GK	56.7	42	12	M	R
JO	75.0	61	16	M	R
PR	64.3	54	12	M	R
DL	80.9	29	16	M	R
TE	64.0	54	14	M	R
HW	77.0	45	16	M	NA
MH	58.0	45	14	F	R
MR	52.3	53	16	F	R
MN	82.1	66	14	F	R
MD	81.5	65	20	M	L
LD	69.6	64	16	F	R
RH	72.0	40	16	M	R
CH	77.2	42	14	M	R
FP	68.	40	14	F	R
MW	NA	49	NA	M	NA
SW	NA	78	14	F	NA
GG	85.2	67	14	M	R
BS	70.6	78	16	F	R

matical elements could be modeled as a single variable, or whether clausal functional categories constituted a separate set. Since plural marking is associated with the determiner/noun phrase rather than the clausal structure, it was predicted that plural morphology would not group with other functional categories. Nouns and verbs were also coded. Note that only semantically and syntactically correct productions were tallied. Criteria used for coding each of these morpho-syntactic elements and examples of coding are provided in [Appendices A and B](#), respectively.

### 2.2.2. Coding reliability

Point-to-point inter-coder reliability was assessed on 30% (6 of 18) of patient transcripts. Mean percent agreement for each target element (complementizers, verb inflections, negation markers, aspectual verb forms, and plurals) was 96%, and ranged from 93% for plurals to 96% for negation markers.

### 2.2.3. Scoring

For each participant the total number of words, clauses, clausal functional categories, plural morphemes, nouns, and verbs were tallied, and the following ratios were computed:

- i. Total words/Total clauses
- ii. Total complementizers/Total clauses
- iii. Total verb inflections/Total clauses
- iv. Total tense inflections/Total clauses
- v. Total agreement inflections/Total clauses
- vi. Total infinitive 'to' particles/Total clauses
- vii. Total modals/Total clauses
- viii. Total negation markers/Total clauses
- ix. Total aspectual verb forms/Total clauses
- x. Total verbs/Total clauses

- xi. Total nouns/Total clauses
- xii. Total plural morphology/Total clauses
- xiii. Total nouns/Total verbs

### 2.2.4. Comparison of control and aphasic data

Group means and standard deviations were computed for both groups on all production measures. *T*-tests for samples with unequal variance were then used to compare performance of the two groups on all production measures.

### 2.2.5. IRT analysis

Means and standard deviations from the control transcripts were used to establish ability/impairment scores for each aphasic participant on each morpho-syntactic element (clausal functional categories and plural morphology). If the morpheme to clause ratio for an aphasic participant was within two standard deviations of the mean for controls, a score of '1' was assigned, if the ratio was more than 2 standard deviations below normal a score of '0' was assigned. Thus, the performance of each aphasic participant was summarized as a row of data containing 0's and/or 1's (as in [Table 1](#)). These data were then analyzed using Winsteps Rasch IRT software ([Linacre & Wright, 2000](#)). The following IRT values were examined: item (functional category) difficulty, person ability, standard error of measurement (*SE*) for both item difficulty and person ability, and model fit (*OMSE*) for item difficulty and person ability. By convention, individuals/items with *OMSE* values greater than or equal to 1.5 can be interpreted as having poor fit with the model ([Bond & Fox, 2001](#)). An overview of the numeric derivation of each value is provided in [Appendix C](#). A more detailed discussion of the derivation of these values can be found in [Wright and Stone \(1979\)](#).

### 2.2.6. Correlation analysis

Pearson  $r$  correlation coefficients were computed to explore the relation between functional category production and other aspects of language processing. First, to assess the relation between functional category production and general language processing, IRT scores were correlated with the WAB AQ. Second, to examine the relation between functional category production and general grammatical processing, IRT scores were correlated with the N/V ratio. Lastly, to explore the relation between functional category production and an independent measure of clausal structure, IRT scores were correlated with clause length (words per clause).

## 3. Results

First, group data are presented comparing the relative frequency with which functional categories were produced by control and aphasic participants. Results of the IRT analysis are then presented. Lastly, correlation data comparing IRT derived ability scores to more general language measures are provided.

### 3.1. Group analysis

Means and standard deviations for the aphasic and control groups on each of the production measures are shown in Table 3. As expected, all control narratives included complementizers, verb inflections, negation markers, aspectual markers, and plural inflections. The different morpheme types varied, however, in terms of the frequency with which they were produced in a clause. Verb inflections were produced with the highest frequency (mean = .89, i.e., 89% of clauses included verb inflections), followed by plural inflections (mean = .23), participle verb forms (mean = .18), complementizers (mean = .11), and negation markers (mean = .06). This pattern was generally stable across participants (see Fig. 4), with the following exceptions: 3 sub-

Table 3  
Comparison of control and aphasic participants on production measures

Language variable	Control participants ( $n = 18$ ) Mean (SD)	Aphasic participants ( $n = 18$ ) Mean (SD)
Total clauses	75.56 (47.81)	*31.33 (15.36)
Words/Clause	6.06 (.41)	*4.00 (.82)
Complementizers/Clause	.11 (.04)	*.03 (.04)
Verb inflections/Clause	.89 (.06)	*.65 (.33)
Tense Infl./Clause	.50 (.29)	.39 (.30)
Agreement Infl./Clause	.21 (.26)	.18 (.20)
“to”/Clause	.14 (.04)	*.06 (.07)
Modal/Clause	.05 (.04)	*.02 (.03)
Negation markers/Clause	.06 (.03)	.04 (.04)
Aspectual markers/Clause	.18 (.08)	.16 (.12)
Verbs/Clause	1.19 (.07)	*1.01 (.19)
Nouns/Clause	1.05 (.13)	*.74 (.25)
Plural inflections/Clause	.23 (.05)	*.13 (.10)
Noun/Verb	.88 (.13)	.77 (.32)

\* Statistically significant based on two sample  $t$ -test (unequal variance)  $p < .01$ .

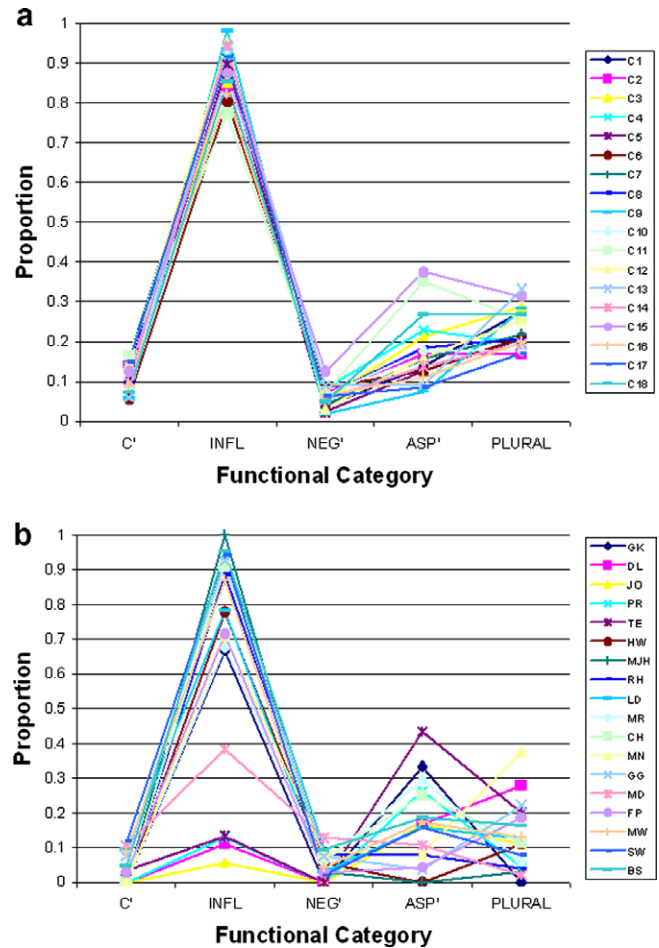


Fig. 4. Variability in the production of functional categories in English clauses. (a) Control participants ( $n = 18$ ). (b) Aphasic participants ( $n = 18$ ).

jects produced more aspectual verb forms than plurals, 1 subject produced more complementizers than aspectual verb forms, and 2 subjects produced more negation markers than complementizers. There was more variability, however, across control participants in the production of particular verb inflections. While the majority of controls produced a variety of inflections, 4 participants omitted agreement inflections, 1 participant omitted past tense morphology, and two participants omitted modals.

Aphasic participants had lower mean scores on all morpho-syntactic production measures, supporting their diagnostic classification (see Table 3). The production frequency for different morpho-syntactic elements differed from controls in several respects: aspectual verb forms were produced with greater frequency than plural inflections, negation markers were produced with greater frequency than complementizers, and importantly there was greater variability observed in the performance patterns of individual patients across all elements (see Fig. 4b).  $T$ -tests for two samples with unequal variance were performed to determine if differences on production measures for the two subject groups were statistically significant (see Table 3). Results indicated that aphasic participants produced signif-

icantly fewer clauses per narrative, words per clause, complementizers per clause, total verb inflections per clause, verbs per clause, nouns per clause, and plural inflections per clause. Differences in the number of negation markers and aspectual verb forms, however, were not significant.

### 3.2. IRT analysis

An IRT analysis was conducted to further evaluate the performance patterns of individual aphasic participants and investigate whether a consistent pattern of impairment was evident across functional categories. To begin, performance of each subject on each functional category was summarized as a single row of data (see Table 4). As described above, a score of '1' was used to indicate preserved performance (within 2 standard deviations of controls) and a score of '0' was used to indicate impaired performance (more than 2 standard deviations below controls). A hierarchical ordering of impairment (complementizers > verb inflections > negation markers > aspectual verb forms) was observed for most subjects across the majority of clausal functional categories. Nonetheless, some variability was observed. For example, participant 4 demonstrated reduced production of aspectual verb forms but produced negation markers within the normal range. Similarly, participant 8 produced both verb inflections and negation markers within the normal range, while production of aspectual verb forms was reduced. In addition, participant 13 produced complementizers within the normal range, but was impaired in his production of verb inflections. Note, however, that performance on plural inflections did not fit the general pattern observed for clausal functional categories. For instance, participants 5 and 6, in spite of demonstrating relatively low general ability

scores ( $\beta = 2$ ) were able to produce plural 's' within the normal range, while participants 14, 15, and 16 had higher overall scores ( $\beta = 4$ ) but failed to produce the normal frequency of plurals in their narratives.

The raw data shown in Table 4 were entered into the Rasch Analysis to evaluate (1) whether clausal functional categories (items) can be grouped together as a single variable fitting a hierarchical model, and (2) to determine the relative order of impairment across functional categories. IRT values for functional categories (model fit, functional category difficulty, measurement error) and participants (model fit, participant ability, measurement error) are shown in Tables 5 and 6, respectively.

#### 3.2.1. Functional category (item) data

Model fit (OMSE), difficulty scores ( $d$ ) and associated measurement error ( $SE_d$ ) for each of the morpho-syntactic elements (clausal functional categories and plural morpheme) are shown in Table 5. As described in Section 2, OMSE was used to assess model fit. By convention, any item with an OMSE greater than 1.5 (50% more variability than predicted) is interpreted as not fitting the model. As shown in Table 5, OMSE values for clausal functional categories (complementizers, verb inflections, negation markers, and aspectual verb forms) ranged from .4 to .82 and therefore can be interpreted as demonstrating good fit with the model and patterning as a coherent set. The plural morpheme was the only element with an OMSE value greater than 1.5 (OMSE = 1.69), indicating that production of this morpheme did not fit the model.

The IRT functional category (item) difficulty scores ( $d$ ) are shown in the third column of Table 5. The order of item difficulty was as follows: complementizers and plurals ( $d = 1.57$ ) > verb inflections ( $d = .76$ ) > negation markers ( $d = -1.17$ ) > aspectual verb forms ( $d = -2.73$ ). The measurement error for item difficulty scores ( $SE_d$ ) ranged from 0.66 to 0.83, indicating that all differences in item difficulty were statistically significant.

#### 3.2.2. Participant data

Table 6 shows model fit (OMSE), ability scores ( $b$ ) and measurement error ( $SE_b$ ) for each of the aphasic participants. Fifteen of 18 participants (83%) had OMSE values below 1.5 and hence were judged to fit the model. Three participants (MH, DL, and TE), however, had OMSE values greater than 1.5. As shown in Table 4, the response pattern of MH was unusual in that production of aspectual

Table 4  
Patterns of functional category production: aphasic participants

Aphasic participant	C'	INFL'	NEG'	ASP'	Plural	Person ability score ( $\beta$ )
1. GK	0	0	0	1	0	1
2. JO	0	0	0	1	0	1
3. PR	0	0	0	1	0	1
4. HW	0	0	1	0	0	1
5. DL	0	0	0	1	1	2
6. TE	0	0	0	1	1	2
7. MR	0	0	1	1	0	2
8. MH	0	1	1	0	0	2
9. FP	0	0	1	1	1	3
10. MN	0	0	1	1	1	3
11. RH	0	1	1	1	0	3
12. CH	0	1	1	1	0	3
13. MD	1	0	1	1	0	3
14. MW	1	1	1	1	0	4
15. SW	1	1	1	1	0	4
16. LD	1	1	1	1	0	4
17. GG	1	1	1	1	1	5
18. BS	1	1	1	1	1	5
Functional category difficulty score ( $\delta$ )	6	8	13	16	6	

Table 5  
IRT analysis of functional categories (items)

Functional category	Model fit (OMSE)	Functional category difficulty score ( $d$ )	Measurement error ( $SE_d$ )
C'	0.40	1.57	.66
INFL'	0.65	0.76	.62
NEG'	0.59	-1.17	.66
ASP'	0.82	-2.73	.83
Plural	1.69*	1.57	.66

\* OMSE > 1.5 (rejected from model).

Table 6  
IRT analysis of participant ability

Aphasic participant	Model fit (OMSE)	Participant ability score ( <i>b</i> )	Measurement error ( <i>SE<sub>b</sub></i> )
1. GK	0.21	-2.17	1.38
2. JO	0.21	-2.17	1.38
3. PR	0.21	-2.17	1.38
4. HW	0.21	-2.17	1.38
5. DL	2.09*	-0.51	1.21
6. TE	2.09*	-0.51	1.21
7. MR	0.23	-0.51	1.21
8. MH	2.70*	-0.51	1.21
9. FP	0.76	0.80	1.11
10. MN	0.76	0.80	1.11
11. RH	0.41	0.80	1.11
12. CH	0.41	0.80	1.11
13. MD	0.76	0.80	1.11
14. MW	0.52	2.08	1.21
15. SW	0.52	2.08	1.21
16. LD	0.52	2.08	1.21
17. GG	MAX <sup>†</sup>	3.58	1.92
18. BS	MAX <sup>†</sup>	3.58	1.92

\* OMSE > 1.5 (rejected from model).

<sup>†</sup> No variability in scores therefore no model error provided.

verb forms was reduced while production of both negation markings and inflections were within the normal range. The response patterns of DL and TE were remarkable in that both participants were able to produce a normal quantity of plural morphemes (a difficult item) in spite of their relatively weak overall performance. In fact, when the analysis was re-run with only clausal functional categories (and the misfitting plural morpheme excluded) only one participant (MH) remains as an outlier.

As shown in Table 6, IRT participant ability scores (*b*) ranged from -2.17 to 3.58, indicating that participants demonstrated a wide range of ability in their production of clausal functional categories. The full IRT model, showing both participant ability and functional category difficulty is illustrated in Fig. 5.

### 3.3. Correlation analyses

To examine the relation between functional category production and other language measures, the IRT participant ability scores were correlated with clause length (words/clause), the N/V ratio, and WAB A.Q. The IRT scores showed a modest correlation with clause length ( $r = .55, p = .02$ ), but were not correlated with the N/V ratio ( $r = -.08, p = .76$ ) or with the WAB A.Q. ( $r = .38, p = .14$ ).

## 4. Discussion

In this paper we adopted a psychometric approach and used IRT to investigate three related questions: (1) Can clausal functional categories be grouped together as a coherent set of items defining a single variable?; (2) Is the order of impairment across functional categories consistent with the hierarchical structure of syntactic trees; and (3) Is there a relation between clausal functional category production

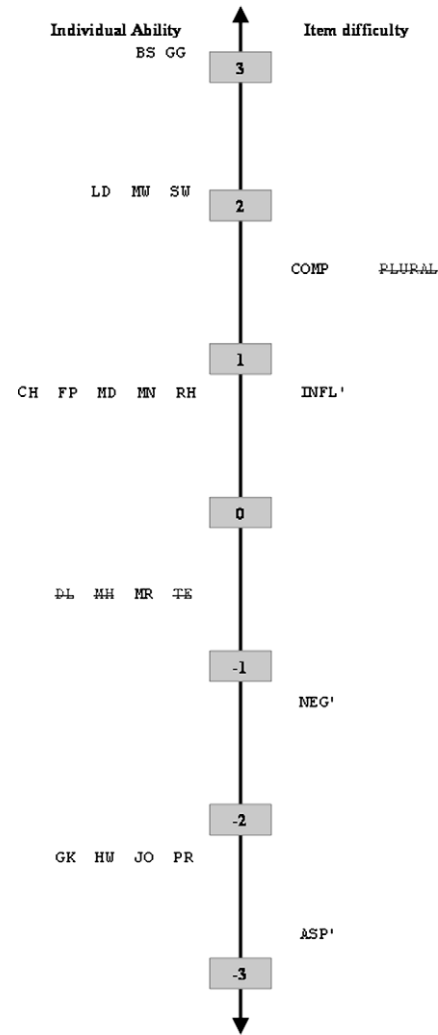


Fig. 5. IRT logit scale showing participant ability and item (functional category) difficulty. Note: Misfitting items and participants are crossed out and shown in gray font.

and performance on other language measures? In the discussion that follows we summarize results of this study with respect to each of these questions, and further interpret these findings in light of on going theoretical debate regarding the nature of sentence production deficits in agrammatism. We conclude by discussing the value of using IRT and related statistical models to investigate these issues.

One question addressed in this study was whether clausal functional categories can be represented as items defining a variable within the probabilistic framework of IRT. Specifically, we were interested in testing whether the relation between functional category difficulty and participant ability was such that it would be possible to accurately predict the probability of individuals with a particular level of ability producing specific functional categories (see Fig. 3). Results of the analysis indicated that the actual production of clausal functional categories (complementizers, verb inflection, negation markers, and aspectual verb forms) demonstrated good fit with the IRT-generated scale (OMSE values ranged from .40 to .82, see Table 5). In fact, the IRT

model of functional category production (see Fig. 5) was able to account for the performance patterns observed in 15 of the 18 aphasic participants. To further explore the status of clausal functional categories as a natural class, we also investigated whether this same model could predict actual production patterns for plural 's' morpheme. While the plural morpheme is also a closed class grammatical morpheme, it is typically associated with the noun phrase (or determiner phrase) rather than clausal structure (Ouhalla, 1991). In contrast to the good fit exhibited by the clausal functional categories, the plural morpheme did not fit the model (OMSE = 1.69). Furthermore when data pertaining to production of plural morphology were removed, and the IRT analysis was re-run only with data from production of clausal functional categories (complementizers, verb inflections, negation markers, and aspectual verb forms), the adjusted model was able to account for the response patterns for all but one of the participants. The fact that production patterns for clausal functional categories can be differentiated from the production patterns associated with other types of grammatical morphology (such as the plural morpheme) provides evidence that clausal functional categories constitute a natural class. In addition, the results of this study suggest that production of clausal functional categories can be represented as a series of items defining an IRT-type variable. These findings are consistent with a wide range of research suggesting that clausal functional categories constitute a structured subset of grammatical or closed class morphology (Chomsky, 1995; Friedmann, 2001; Ouhalla, 1991).

A second goal was to evaluate whether functional category production patterns reflected the hierarchical structure of syntactic trees. As shown numerically in Table 5 and graphically in Fig. 5, the order of functional category difficulty was consistent with the hierarchical structure of the syntactic tree: complementizers ( $d = 1.57, \pm .66$ ) > verb inflections ( $d = 0.76, \pm .62$ ) > negation markers ( $d = -1.17, \pm .66$ ) > aspectual verb forms ( $d = -2.73, \pm .83$ ). Note, however, that considerable variability was observed in the production patterns of individual participants. As mentioned above, the IRT-generated scale could not account for the response pattern of MH who demonstrated poor fit with the model. Furthermore, more subtle deviations to expected performance patterns were observed in the production patterns of HW, who produced a normal proportion of negation markers but not aspectual verb forms; and MD, who produced complementizers but not verb inflections. Taken together, these results suggest that the hierarchical order of syntactic trees characterizes the production patterns for the majority of participants on most functional categories. However, this relation is not deterministic; evidence of impairment at a particular structural level does *not necessarily* entail impairment at all higher levels, rather, it *increases the probability* that such impairments will also be evident.

Although some of the variability observed in the production patterns may be attributed to task-specific effects,

particularly the loosely constrained nature of narrative elicitation, several pieces of evidence suggest that the observed variability is unlikely to be explained solely (or primarily) by use of this task. First, this variability was *not* observed in the control participants. As shown in Fig. 4a, the relative frequency with which various functional categories were produced was quite consistent across controls. For instance, note that all control participants produced finite verb inflections more frequently than other functional categories. This contrasts with the performance of the agrammatic speakers (Fig. 4b). Although finite verb inflections were the most frequently produced functional category for some agrammatic speakers, others produced plurals and/or aspectual markers with greater frequency. Given the stable performance patterns observed in the controls, it seems reasonable to infer that the variability observed in patients is closely tied to their impairment rather than task-specific effects. Second, one of the aphasic participants (MD) who demonstrated an atypical pattern (normal production of complementizers but impaired production of verb inflections) has shown this same deviant pattern on a more constrained sentence elicitation task (Lee et al., 2005). Moreover, our results are consistent with the published research on this topic in which an impairment pattern consistent with hierarchical syntactic structure is observed in many (Friedmann & Grodzinsky, 1997; Hagiwara, 1995) but *not* all (Burchert et al., 2005a; Lee, 2003; Lee et al., 2005) cases of agrammatism.

A third objective of this research was to examine the relation between functional category production and performance on other language measures. Production of functional categories (see Table 6 Participant Ability scores) showed a modest correlation with clause length ( $r^2 = .55, p = .02$ ) but was not correlated with the N/V ratio ( $r^2 = -.079, p = .78$ ) or with the WAB AQ ( $r^2 = .38, p = .14$ ). In general, the weak relation observed between functional category production and other language measures suggests that neither the general language performance nor the sentence production deficits in this group of participants can be explained solely in reference to functional category production. This finding is consistent with the view that multiple variables contribute to the sentence production and overall language performance observed in agrammatic aphasia (Caplan, 1991; Caramazza & Berndt, 1985; Goodglass & Menn, 1985). Although this position continues to be debated with respect to agrammatism, it is much more widely accepted in aphasia research focusing on *lexical* deficits (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Shapiro & Caramazza, 2001). Given the weight of clinical (Bastiaanse & Thompson, 2003; Goodglass & Menn, 1985; Rochon et al., 2000; Saffran et al., 1989; Thompson, Shapiro, Kiran, & Sobecks, 2003), psycholinguistic (Dell, Chank, & Griffin, 2001; Garrett, 2000), and linguistic (Chomsky, 1986, 1995) evidence pointing to the complexity of sentence processing, it would be surprising if multiple variables were not also responsible for the diverse production patterns observed in agrammatism.

## 5. Conclusions

In this paper we used IRT to examine production of clausal functional categories in agrammatic narratives. Results of our analyses indicate that functional categories can be represented as a set of items defining a single variable on an IRT-generated scale. Furthermore, for many, but not all, agrammatic individuals the observed production patterns mirrored the hierarchical structure of syntactic trees. Nonetheless, several individuals deviated from the expected pattern, suggesting that this relation is probabilistic in nature rather than absolute. In addition, our results indicate that production of functional categories does not account for the full range of sentence production impairments. The fact that functional category production showed at best a modest correlation with other measures of sentence production (such as clause length, and N/V ratio) supports the view that additional variables are contributing to the language performance patterns observed in this group of participants.

The findings summarized above suggest that IRT may prove to be a useful model to further investigate functional category production and other variables associated with agrammatism. The IRT approach complements other psychometric techniques in several ways. One useful feature of the IRT model is that it makes it possible to test whether clinical/psycholinguistic constructs can be represented as a hierarchically ordered parametric variable. For constructs (such as functional category production) that can be represented as a numeric variable, the resulting equal interval logit scale (see Fig. 5 and Table 5) can be used to generate precise and falsifiable predictions about the nature of the construct. For instance, the functional category difficulty scores (see Table 5) indicate that the magnitude of the difference between complementizers and verb inflections ( $1.57 - 0.76 = .81$  logit units) is smaller than the difference between verb inflections and negation markers ( $0.76 - -1.17 = 1.93$  logit units).

A related advantage of the IRT scale is that it is based on a probabilistic model. The use of probabilities facilitates differentiation of response patterns such as those demonstrated by DL, MH, and TE (which depart significantly from model expectations) and those demonstrated by HW and MD (which are consistent with the probabilistic model's expected variability). An important goal for future research is to determine whether the properties of the functional category scale shown in Fig. 5 can be replicated with data collected using the same task (narrative elicitation) as well as with data collected using more constrained sentence elicitation procedures.

Another unique property of the IRT scale is that items and individuals are measured on the same scale. Consequently, the same model can be used to identify items and individuals that are statistical outliers. For example, in the functional category scale, the 'plural' item was identified as an outlier, as were three participants who similarly did not fit the model. With respect to the misfitting plural item, note that if we were simply to look at the difficulty scores, we

might be inclined to group complementizers and plural “\_s” together. However, when we examine the IRT fit statistics (how individuals of particular level of ability perform on specific functional categories) it is apparent that complementizers and plural “\_s” morphology can be differentiated. Stated differently, the IRT model provides statistical evidence that participant ability on *clausal* functional category production does not predict performance on DP (plural “\_s”) related morphology. A possible direction for future research would be to explore whether a second IRT scale could be defined for DP related morphology, such as determiners, plural ‘\_s’, and possessive morphology.

In summary, results of this study support the use of IRT as a psychometric tool that can be used to specify functional category production and possibly other variables associated with agrammatism. We are not proposing that exhaustive specification of a single variable, such as clausal functional categories, will lead to a full understanding of agrammatic sentence production patterns, but rather that detailed specification of individual component variables can only enhance our construction of more complex and valid multi-factorial models.

## Appendix A. Criteria for coding grammatical elements

Measure	Coding criteria
C'	Subordinating conjunctions introducing a dependent clause. Examples: The fairy tells her <u>that</u> she better be home by midnight. She had to leave <u>otherwise</u> the carriage would turn into a pumpkin. The courier went into the house <u>that</u> housed Cinderella.
Infinitival marker	Use of 'to' to indicate nonfinite form. Examples: She had the dishes <u>to</u> do. Wants <u>to</u> go a ball. Funny ugly lady tried <u>to</u> put foot in there.
Modals	Modal verbs Examples: can, will, should, could, must
T'	Regular and irregular past tense inflections Examples: <u>ed</u> , was, went, did, etc.
Agr'	Regular and irregular 3rd person agreement inflections Examples: <u>s</u> , is, are, has
Neg'	Negation markers Examples: not , never
Plural	Regular and irregular forms Examples <u>s</u> , mice, men
Nouns	All common and proper nouns Examples: Cinderella, prince, ball, slipper, pumpkin
Verbs	All lexical and auxiliary verbs in base form Examples: sew, dance, run, be, have

## Appendix B. Sample coding (excerpts from six participants)

	Word count	Clausal functional categories					Plural	Nouns	Verbs	
		COMP	INFL			NEG				ASP
			'to'	M	T					
<i>Participant DL</i>										
Cinderella (OK) first (uh) scrub the floor	5						Cinderella, floor	Scrub		
Do the dishes	3					_s	Dish	Do		
(OK and then) Sewing	1					_ing		Sew		
<i>Participant LD</i>										
(step) Stepmother not like her dress	5				not		Stepmother, dress	Like		
So she rip apart	4							Rip		
(and Cinderella and) She was upset	3			was				Be		
<i>Participant RH</i>										
(and uh) Lose the slipper	3						Slipper	Lose		
(and uh) Prince picked it up	4			_ed			Prince	Pick		
(and uh) Woman tried it on	3			_ed			Woman	Try		
Never worked	2			_ed	never			Work		
<i>Participant CH</i>										
It will not fit anybody	5		will		not			Fit		
Finally she has her (uh, sh) glass slipper	6				has		Slipper	have		
(She) Her mother (uh) locked up and	5			_ed			Mother	Lock		
Then she got it	4			got				Get		
<i>Participant FP</i>										
Just try to get the shoes on the girls and	10	to				_s, _s	Shoe, girl	Try, get		
Couldn't	2		could		n't					
mother (was) tried to help too	5	to		_ed			Mother	Try, help		
<i>Participant GG</i>										
The courier (uh) too (uh) went into (the uh) the house	7			went			Courier, house	Go		
(uh) that housed Cinderella	3	that		_ed			Cinderella	House		
She tries the slipper on	5				_s		Slipper	Try		
(It uh) it (uh) fit(ted)	2							Fit		

Note. M = Modal, T = tense inflection (verb), A = agreement inflection (verb). All material in parenthesis, including, fillers ("and uh"), false starts ("she" before "her mother"), and obvious syntactic/semantic violations ("ed" after "fit" → "fitted") were excluded from the analysis.

## Appendix C. Computation of item difficulty score, individual ability score, and model fit

### C.1. Item difficulty

The IRT item difficulty score ( $d_i$ ) is given by Eq. (1):

$$d_i = M_\beta + (1 + \sigma_\beta^2/2.89)^{1/2} \ln[(1 - p_i)/p_i] \quad (1)$$

where ‘ $M_\beta$ ’ and ‘ $\sigma_\beta$ ’ are the mean and standard deviation for the ‘Ability Scores’ of the calibrating sample (see Table 1, far right column); and ‘ $p$ ’ is the proportion of individuals in the sample who answered the specified item correctly (see Table 1, bottom row). As is evident from Eq. (1) the IRT item difficulty score is based on the traditional ‘ $p$ ’ value used in classic measurement theory, which is adjusted to compensate for the ability level of the sample, and transformed using the natural log (ln) to fit an equal interval scale. The standard error of measurement for  $d_i$  ( $SE_{di}$ ):

$$(SE_{di}) = (1 + \sigma_\beta^2/2.89)^{1/2} [1/N p_i(1 - p_i)]^{1/2} \quad (2)$$

is similarly based on the standard error of measurement for the traditional  $p$  value ( $SEp_i = [p_i\{1 - p_i\}/N^{1/2}]$ ), with adjustments made for the variance ( $\sigma_\beta$ ) in the calibrating sample ( $N$ ) and the natural log model.

### C.2. Individual ability

The IRT individual ability score ( $b_v$ ) is shown in Eq. (3):

$$b_v = M_\delta + (1 + \sigma_\delta^2/2.89)^{1/2} \ln[r_v/(L - r_v)] \quad (3)$$

where ‘ $M_\delta$ ’ and ‘ $\sigma_\delta$ ’ are the mean and standard deviation for the Item Difficulty scores (Table 1, bottom row);  $r_v$  is the ability score of person  $v$  (Table 1, far right column);  $L$  is the number of items in the test, and 2.89 is a scaling factor ( $1.7^2$ ) used to create an ogive function. Again, the IRT individual ability score ( $b$ ) is derived from the individual ability (or test) score used in classic measurement theory ( $\beta$ ), and adjusted to compensate for the difficulty level of the items, and transformed using the natural log (ln) to fit an equal interval scale. The standard error of measurement for  $b_v$  ( $SE_{bv}$ ):

$$(SE_{bv}) = (1 + \sigma_\delta^2/2.89)^{1/2} [L/r_v(1 - r_v)]^{1/2} \quad (4)$$

is similarly based on the standard error of measurement for the traditional ability score with adjustments made for the variance in the difficulty of test items ( $\sigma_\delta$ ) and the natural log model.

### C.3. Model fit

Outfit mean square error (OMSE) summarizes the consistency with which individuals and items fit the predicted model. It is based on the difference between the *actual* response (0 or 1) of individual $_v$  on item $_i$  and the *expected* response ( $\beta - \delta$ ) of individual $_v$  on item $_i$ . This difference is

the residual error ( $Z_{vi}$ ). The outfit mean square error for an item (OMSE $_i$ ) is computed by summing the squared standardized residuals of each individual for the specified item and dividing by the total number of individuals ( $N$ ):

$$OMSE_i = \Sigma Z_{vi}^2/N \quad (5)$$

Similarly, the outfit mean square error for an individual (OMSE $_v$ ) is computed by summing the squared standardized residuals of each item for the specified individual and dividing by the total number of items ( $L$ ):

$$OMSE_v = \Sigma Z_{vi}^2/L \quad (6)$$

OMSE can be interpreted as a chi-square statistic (Bond & Fox, 2001). Items and/or individuals with OMSE values approximating ‘1’ demonstrate good fit with the model. The extent to which an OMSE value deviates from ‘1’ provides a measure of model misfit. For example, an OMSE of 1.5 indicates that there is 50% more variability in the observed data than was predicted by the model.

## References

- Abney, S. (1987). *The English noun phrase in its sentential aspects*. Unpublished doctoral dissertation. Cambridge, MA: MIT.
- Arabatzi, M., & Edwards, S. (2002). Tense and syntactic processes in agrammatical speech. *Brain and Language*, 80, 314–327.
- Badecker, W., & Caramazza, A. (1985). On considerations of method and theory governing the use of clinical categories in neurolinguistics and cognitive neuropsychology: The case against agrammatism. *Cognition*, 20, 97–125.
- Bastiaanse, R., & Thompson, C. K. (2003). Verb and auxiliary movement in agrammatical Broca’s aphasia. *Brain and Language*, 84, 286–305.
- Bates, E., McDonald, J., MacWhinney, B., & Appelbaum, M. (1991). A maximum likelihood procedure for the analysis of group and individual data in aphasia research. *Brain and Language*, 40, 231–265.
- Benedet, M. J., Christiansen, J. A., & Goodglass, H. (1998). A cross-linguistic study of grammatical morphology in Spanish- and English-speaking agrammatical patients. *Cortex*, 34, 309–336.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Briel, J. B., O’Neill, K. A., & Scheuneman, J. D. (1993). *GRE technical manual: Test development, score interpretation, and research for the Graduate Record Examinations Program*. Princeton, NJ: Educational Testing Service.
- Burchert, F., Swoboda-Moll, M., & De Bleser, R. (2005a). Tense and Agreement dissociations in German agrammatical speakers: Underspecification vs. hierarchy. *Brain and Language*, 94, 188–199.
- Burchert, F., Swoboda-Moll, M., & De Bleser, R. (2005b). The left periphery in agrammatical clausal representations: Evidence from German. *Journal of Neurolinguistics*, 18, 67–88.
- Caplan, D. (1991). Agrammatism is a theoretically coherent aphasic category. *Brain and Language*, 40, 274–281.
- Caramazza, A., & Berndt, R. S. (1985). A multicomponent deficit view of agrammatical Broca’s aphasia. In M.-L. Kean (Ed.), *Agrammatism*. San Diego: Academic Press.
- Chomsky, N. (1986). *Barriers*. Cambridge: MIT Press.
- Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.
- Cinque, G. (2002). *Functional structure in DP and IP*. New York: Oxford University Press.
- Dell, G. S., Chan, F., & Griffin, Z. M. (2001). Connectionist models of language production: Lexical access and grammatical encoding. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics*. Westport, CT: Ablex Publishing.

- Dell, G. S., Schwartz, M. F., Martin, N. M., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801–838.
- Doyle, P. J., McNeil, M. R., Mikolic, J. M., Prieto, L., Hula, W. D., Lustig, A. P., et al. (2004). The Burden of Stroke Scale (BOSS) provides valid and reliable score estimates of functioning and well-being in stroke survivors with and without communication disorders. *Journal of Clinical Epidemiology*, *57*(10), 997–1007.
- Drai, D., & Grodzinsky, Y. (2006). A new empirical angle on the variability debate: Quantitative neurosyntactic analyses of a large data set from Broca's Aphasia. *Brain and Language*, *96*, 117–128.
- Friedmann, N. (2001). Agrammatism and the psychological reality of the syntactic tree. *Journal of Psycholinguistic Research*, *30*, 71–88.
- Friedmann, N. (2002). Question production in agrammatism: The tree pruning hypothesis. *Brain and Language*, *80*, 160–187.
- Friedmann, N., & Grodzinsky, Y. (1997). Tense and agreement in agrammatic production: pruning the syntactic tree. *Brain and Language*, *56*, 397–425.
- Garrett, M. (2000). Remarks on the architecture of language processing systems. In Y. Grodzinsky, L. P. Shapiro, & D. Swinney (Eds.), *Language and the brain: Representation and processing*. San Diego: Academic Press.
- Goldstein, K. (1948). *Language and language disturbances*. New York: Grune & Stratton.
- Goodglass, H., & Berko, J. (1960). Agrammatism and inflectional morphology in English. *Journal of speech and hearing research*, *3*, 257–269.
- Goodglass, H., & Menn, L. (1985). Is agrammatism a unitary phenomenon?. In M.-L. Kean (Ed.) *Agrammatism*. San Diego: Academic Press.
- Granger, C. V., Hamilton, B. B., Linacre, J. M., Heinemann, A. W., & Wright, B. D. (1993). Performance profiles of the functional independence measure. *American Journal of Physical Medicine & Rehabilitation*, *72*(2), 84–89.
- Haegeman, L. (1994). *Introduction to government and binding theory* (2nd ed.). Cambridge, MA: Blackwell.
- Hagiwara, H. (1995). The breakdown of functional categories and the economy of derivation. *Brain and Language*, *50*, 92–116.
- Hula, W., Doyle, P. J., McNeil, M. R., & Mikolic, J. M. (2006). Rasch modeling of Revised Token Test performance: Validity and sensitivity to change. *Journal of Speech, Language, and Hearing Research*, *49*, 27–46.
- Kertesz, A. (1982). *Western aphasia battery*. San Antonio: Psychological Corporation.
- Koenig, J.A. (1998). *MCAT interpretive manual: A guide for understanding and using MCAT scores in admissions decisions*. Washington, DC: AAMC.
- Lee, M. (2003). Dissociations among functional categories in Korean agrammatism. *Brain and Language*, *84*, 170–188.
- Lee, J., Milman, L. H., & Thompson, C. K. (2005). Functional category production in agrammatic speech. *Brain and Language*, *95*, 123–124.
- Linacre, J. M., & Wright, B. D. (2000). *Winsteps: Multiple-choice, rating scale, and partial credit Rasch analysis*. Chicago: MESA Press, [computer software].
- McCloskey, M. (1993). Theory and evidence in Cognitive Neuropsychology: A 'radical' response to Robertson, Knight, Rafal, and Shimamura (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 718–734.
- Menn, L., & Obler, L. K. (Eds.). (1990). *Agrammatic aphasia: A cross-language narrative sourcebook*. Philadelphia: John Benjamins Publishing.
- 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., 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.
- Milman, L. H. (1997). *Processing inflectional verb morphology in Spanish: Evidence from Spanish agrammatism*. Unpublished Master's Thesis, University of Arizona, Tucson.
- Milman, L. H., & Holland, A. (November, 2004). *The SCCAN: A new test measuring adult cognition and communication*. Paper presented to the American Speech Language Hearing Convention. Philadelphia, PA.
- Myerson, R., & Goodglass, H. (1972). Transformational grammars of three agrammatic patients. *Language and Speech*, *15*, 40–50.
- Ouhalla, J. (1991). *Functional categories and parametric variation*. New York: Routledge.
- Ouhalla, J. (1993). Functional categories, agrammatism and language acquisition. *Linguistische Berichte*, *143*, 3–36.
- Pollock, J. Y. (1989). Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry*, *20*, 365–424.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago: University of Chicago Press.
- Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and Language*, *72*, 193–218.
- Saffran, E., Berndt, R. S., & Schwartz, M. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language*, *37*, 440–479.
- Shapiro, K., & Caramazza, A. (2001). Language is more than its parts: A reply to Bird, Howard, & Franklin. *Brain and Language*, *78*, 397–401.
- Stavrakaki, S., & Kouvava, S. (2003). Functional categories in agrammatism: Evidence from Greek. *Brain and Language*, *86*, 129–141.
- Thompson, C. K., Fix, S., & Gitelman, D. (2002). Selective impairment of morphosyntactic production in a neurological patient. *Journal of Neurolinguistics*, *15*, 189–207.
- Thompson, C. K., Shapiro, L. P., Kiran, S., & Sobecks, J. (2003). The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of treatment efficacy (CATE). *Journal of Speech, Language, and Hearing Research*, *46*(3), 591–607.
- Thompson, C. K., Shapiro, L. P., Tait, M. E., Jacobs, B. J., Schneider, S. L., & Ballard, K. J. (1995). A system for the linguistic analysis of agrammatic language production. *Brain and Language*, *51*, 124–129.
- Wechsler, D. (1997a). *Wechsler adult intelligence scale-III*. San Antonio: Psychological Corporation.
- Wechsler, D. (1997b). *Wechsler memory scale-III*. San Antonio: Psychological Corporation.
- Wenzlaff, M., & Clahsen, H. (2005). Finiteness and verb-second in German agrammatism. *Brain and Language*, *92*, 33–44.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.