Functional categories related to verb inflection are not differentially impaired in Greek aphasia*

ATHANASSIOS PROTOPAPAS

Department of Philosophy & History of Science, University of Athens, Greece

SPYRIDOULA CHEIMARIOU

Graduate Program in Basic and Applied Cognitive Science, University of Athens, Greece

ALEXANDRA ECONOMOU

Department of Psychology, University of Athens, Greece

MARIA KAKAVOULIA

Department of Media, Communication, and Culture, Panteion University, Athens, Greece

AND

SPYRIDOULA VARLOKOSTA

Department of Linguistics, Faculty of Philology, University of Athens, Greece

ABSTRACT

Previous research in Greek aphasia has indicated that functional categories related to verb inflection are differentially impacted, with Aspect most

severely affected, Agreement least affected, and Tense occupying an intermediate position. However, research materials were not controlled for overall length or position of the verb within the sentence, confounding functional category with processing load. Using balanced materials, here we tested ten persons with aphasia and ten matched control participants on grammaticality judgment and sentence completion in three functional categories (agreement, tense, and aspect) using ten verbs spanning a range of familiarity. Production results indicated no difference in errors of either lexeme or inflectional morpheme selection. In grammaticality judgment acceptance of incorrect sentences was lower for Agreement but this pattern was mirrored in the control group as well. The results provide no basis to support a specific linguistic deficit in the representation of functional categories in Greek persons with aphasia.

KEYWORDS: aphasia, Greek, functional categories, verb, processing.

1. Introduction

Non-fluent aphasia is an acquired language disorder often characterized by impoverished syntactic structures and omission or substitution of grammatical morphemes, termed 'agrammatism' (Menn & Obler, 1989). Agrammatism is considered a key disorder for the investigation of language and the convergence of linguistics with neuroscience (Beretta, 2008). Two broad theoretical approaches have been proposed to describe it, namely the structural/representational approach and the processing approach (see Bastiaanse & Thompson, 2012). The former account attributes impairments to the breakdown of grammatical representations in one or more levels of linguistic analysis (e.g., Kean, 1977, 1979). In contrast, the latter account suggests that agrammatism reflects an inability to effectively employ grammatical knowledge due to lack of processing resources (e.g., Kok, van Doorn, & Kolk, 2007).

Selective deficits in specific grammatical phenomena are often regarded as evidence supporting the representational approach, consistent with the existence of specialized grammar modules that can be damaged separately. A long-standing debate concerns deficits in verb morphology. Several neurolinguistic studies have suggested that not all grammatical morphemes are equally disrupted in agrammatism, and a variety of theoretical proposals have been offered to account for the observed asymmetries (see Bastiaanse & Grodzinsky, 2000; Bastiaanse & Thompson, 2012). In Greek agrammatic aphasia, subject–verb agreement generally appears better preserved than tense and aspect (Fyndanis, Varlokosta, & Tsapkini, 2012; Nanousi, Masterson, Druks, & Atkinson, 2006; Varlokosta, Valeonti, Kakavoulia, Lazaridou, Economou, & Protopapas, 2006). More specifically, Varlokosta et al. investigated production and reception of subject–verb agreement, tense, and aspect in Greek

speakers with aphasia, using a sentence completion task and a grammaticality judgment task. Their results indicated selective deficits, in that participants performed relatively poorly in tense and aspect while performance on subject–verb agreement was comparatively higher. Varlokosta et al. proposed an explanation based on the distinction between interpretable and uninterpretable features, which can be construed either as a representational account or as consistent with a processing deficit.

Crucially, performance on language tests is necessarily mediated by cognitive processes for the perception, access, maintenance, manipulation, and production of the required verbal sequences. Barriers to any of these cognitive processes may affect task performance without necessarily being specific to linguistic processing, linguistic structure, or linguistic representations. Differences in the cognitive processing requirements of different functional categories will likely lead to differential deficits in language tasks that engage these categories. Therefore, any parsimonious account of specific linguistic deficits must methodologically exclude simpler, more general accounts that do not require strong theoretical commitment to specific units or structures. That is, linguistic tasks assessing different functional categories must first ensure that processing requirements are comparable among the contrasted categories.

In this respect, the experimental design of Varlokosta et al. (2006) precludes robust conclusions because the materials used in their study were not balanced across the three functional categories. In order to elicit the desired verb forms, different sentence lengths were used in each condition, resulting from inherent differences in the number of critical cues needed to elicit the different functional categories. This was done to minimize load and avoid redundancy, while adequately constraining the phrase context towards the target form. For example, in the agreement condition only the subject needs to be processed to produce the correct form. Therefore, agreement items were simple short sentences including only a subject, a verb, and an object. However, in the tense condition the sentences were augmented by a temporal adverb denoting the time of action, while in the aspect condition a temporal adverb and an adverbial phrase denoting perfective or imperfective aspect were needed to constrain the intended form. Therefore, sentence length was increasing across the three conditions in the order Agreement < Tense < Aspect, the same order in which performance deficits were observed. Although length is not the only difference relevant for processing, differences in sentence length confounded functional category with amount of information to be processed. Thus, an alternative, simpler explanation is conceivable, accounting for the significant performance differences among functional category conditions in terms of processing load rather than linguistic structure.

Sentence length matching among conditions is necessary for a well-controlled study (Bastiaanse, Bouma, & Post, 2009). Therefore, in the present study we aimed to address this weakness by equating testing materials in length

and other properties across the three functional categories, to the extent possible. In addition, we increased the number of verbs and their range of aspectual formation paradigms. Except for these specific enhancements in the materials, we closely followed the method of Varlokosta et al. (2006) in task design, participant recruitment, and testing procedure. We investigated the performance of ten participants with aphasia and a matched control group in agreement, tense, and aspect, in both production and reception, using a sentence completion task and a grammaticality judgment task, respectively. If differences in performance among the three functional categories persisted, they could more safely be interpreted as of specifically linguistic origin. If, on the other hand, no differences among the three functional categories were found, the earlier findings would be consistent with a processing deficit commensurate with the length of the sentence used to elicit the functional category. Therefore, the results of the present study help disambiguate between two possible kinds of explanations for the findings of Varlokosta et al. (2006).

2. Method

2.1. PARTICIPANTS

Ten individuals (one woman), 51–74 years old (M = 61.9, SD = 9.8) with 8–17 years of education (M = 12.2, SD = 2.4), clinically diagnosed with aphasia, participated in the study. All participants were right-handed, monolingual native speakers of Greek. They all had suffered a (unilateral) left hemisphere lesion due to a single cerebrovascular accident 9–61 months prior to testing (M = 21.2, SD = 15.6). Fluency deficits ranged from mild (3 participants) and moderate (6) to severe (1).

All participants with aphasia were clinically determined to be free of dementia. The Boston Diagnostic Aphasia Examination 3rd Edition–Short Form (BDAE-SF; Goodglass, Kaplan, & Barresi, 2001, adapted in Greek by Tsapkini, Vlahou, & Potagas, 2009/2010) was used to assess language deficits. To assess evidence for agrammatism in language production, patients were also tested on a picture description task using the Cookie Theft picture from the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1983).

Data were collected in compliance with the regulations of the Eginition Hospital ethics committee. Participation was voluntary and participants were informed that they could discontinue testing at any time.

A control group of ten non-impaired participants were recruited, matched with the speakers with aphasia on sex, age (M = 59.9, SD = 9.3), and education (M = 11.6, SD = 3.7). All participants reported normal or corrected to normal vision, normal hearing, and no history of medical or psychiatric illness.

Table 1 lists demographic and clinical information for the participants.

2.2. MATERIALS

Ten transitive, two-syllable verbs, stressed on the penultimate syllable in their base form, were used to construct the sentences; they are listed in Table 2. Four of the verbs formed a regular perfective aspectual theme (rule-based paradigm), four verbs formed a semi-regular theme (stored allomorph paradigm), and two verbs formed an irregular perfective aspectual theme (see Holton, Mackridge, & Philippaki-Warburton, 1997, for information on Greek verb inflection, and Varlokosta et al., 2006; Tsapkini, Jarema, & Kehayia, 2002, for discussion of the different types of past perfective formation in more detail, along the lines of Ralli, 1988).

The verbs were rated for subjective familiarity by twenty adults (not including any of the patient or control participants) on a scale of 1 (low: used 'rarely, if ever') to 5 (high: used 'every day'). There were two highly familiar and two less familiar regular and semi-regular verbs; both irregular verbs were highly familiar. Each regular and irregular pair included one verb with a consonant cluster in the stem and one with no clusters (this was not possible for the semi-regular pairs). Familiarity, regularity, and phonological complexity produced no consistent effects and are not examined in this report.

These verbs were used to construct sentences for both tasks in the three grammatical category conditions. All sentences were affirmative and included one verb in the active voice. Following closely Varlokosta et al. (2006), four target sentences were constructed for each of the ten verbs in each condition, each complemented with a corresponding cue sentence (for sentence completion) and an incorrect sentence (for grammaticality judgment), for a total of 120 test items in the sentence completion task (each target sentence elicited using the corresponding cue sentence) and 240 test items in grammaticality judgment (each target sentence and each incorrect sentence, judged separately).

For the agreement condition, half of the sentences tested number and half tested person. For example, an item testing number in the third person was /polés forés to çimona ta peðjá xánun to proinó tréno/ 'Many times during the winter the children miss$_{3RD.PL}$ the morning train' (cue sentence) || /polés forés to çimona to peðí/ 'Many times during the winter the child _____' (incomplete target sentence, to be continued in the sentence completion task as /xáni to proinó tréno/ 'miss$_{3RD.SG}$ the morning train'). An item testing first person, given third person as a cue, was /símera óli méra o mános γráfi γráma sti θía/ 'All day today Manos writes$_{3RD.SG}$ a letter to aunt' || /símera óli méra eγó/ 'All day today I _____' (to be completed as /γráfo γráma sti θía/ 'write$_{1ST.SG}$

TABLE 1. *Characteristics of participants with aphasia*

Person	Age	Sex	Education (years)	Months post-onset	Lesion site[a]	Fluency deficit[b]
P01	54	M	12	19	IP, ST, MT, IT	moderate
P02	57	M	17	18	Put, PrC, Ins, IF, MF, ST	moderate
P03	70	M	12	25	P, O, BG	severe
P04	56	M	12	18	eC, Put, Ins, IF, IP, ST	moderate
P05	73	M	8	12	Ins, IP	mild
P06	60	M	10	16	iC, Put	mild
P07	74	F	12	61	Put, Cau, PrC, Ins, IF, IP, ST	moderate
P08	74	M	12	4.5	n/a	moderate
P09	50	M	15	29	n/a	moderate
P10	51	M	12	9	iC, eC, GP, Put, IP	mild

NOTES: [a] Based on the radiology report; I = inferior, M = middle, S = superior, F = frontal, P = parietal, T = temporal, O = occipital, PrC = precentral, Put = putamen, Cau = caudate, Ins = insula, e = external, i = internal, C = capsule, BG = basal ganglia, GP = globus pallidus, n/a = data not available. [b] Clinically determined.

a letter to aunt'). For the tense condition, half of the sentences tested the past tense and half the future tense. For example, a test item was /fétos i θía eléni ólo xáni ta jaʎá tis/ 'This year aunt Helen constantly loses her glasses' (cue) || /périsi i θía eléni ólo/ 'Last year aunt Helen constantly _____' (target, completed as /éxane ta jaʎá tis/ 'lost$_{\text{IMPERF.}}$ her glasses'). For the aspect condition, half of the target sentences were in the perfective and half in the imperfective aspect. For example, a test item was /apo ávrio o θános sinéçia θa vlépi ton patéra tu/ 'Starting tomorrow Thanos will constantly be seeing his father' (cue) || /ávrio o θános ksafniká/ 'Tomorrow Thanos suddenly _____' (target, completed as /θa ðí ton patéra tu/ 'will see$_{\text{PERF.}}$ his father'). See Varlokosta et al. (2006) for further details of sentence construction and conditions.

To address the weakness in the experimental design of Varlokosta et al. (2006), materials were balanced across functional categories for total length of phrase. Specifically, sentences were precisely equated in number of characters ($M = 48.0$ in each of the 3 conditions), number of words ($M = 8.6$), and number of words preceding the verb ($M = 4.9$).

2.3. PROCEDURE

The two experimental tasks were administered individually to each participant. The three conditions in each task, addressing participants' performance in the three functional categories, were administered separately, in the same order for all participants (agreement, tense, aspect). Short breaks were taken as required.

TABLE 2. *Properties of the ten verbs used in the test sentences*

Present	Past	Regularity	Subjective familiarity	Estimated frequency	Consonant cluster in stem
ɣrafo 'I write'	εɣrapsa 'I wrote'	regular	4.0	frequent	Yes
plεko 'I weave'	εplεksa 'I wove'	regular	1.4	infrequent	Yes
xano 'I lose'	εxasa 'I lost'	regular	3.4	frequent	No
ðεno 'I tie'	εðεsa 'I tied'	regular	2.5	infrequent	No
vlεpo 'I see'	iða 'I saw'	irregular	4.5	frequent	Yes
lεo 'I say'	ipa 'I said'	irregular	4.4	frequent	No
ðεrno 'I beat'	εðira 'I beat'	semi-regular	1.2	infrequent	Yes
plεno 'I wash'	εplina 'I washed'	semi-regular	3.8	frequent	Yes
sεrno 'I drag'	εsira 'I dragged'	semi-regular	2.0	infrequent	Yes
stεlno 'I send'	εstila 'I sent'	semi-regular	3.2	frequent	Yes

NOTE: Regular verbs retain their stem and add the -*s*- affix; semi-regular verbs undergo a stem vowel change and omit the -*s*- affix; irregular verbs undergo a complete stem change. Subjective familiarity was rated on a scale of 1 (low: used 'rarely, if ever') to 5 (high: used 'every day'). Frequency was estimated on the basis of mean subjective familiarity median split (frequent > 3; infrequent < 3).

In both tasks, cue sentences were presented orally to the participants and the participants' responses were always oral. For sentence completion, participants were told that they were going to hear a sentence and then the first part of a second sentence. They were instructed to complete the second sentence with the appropriate modification. Two or more examples were presented until it was clear that the participant responded appropriately. For the grammaticality judgment task, each participant was asked to judge the grammaticality of each sentence. There was no time limit for responses in either task, though the experimenter moved on to the next item if there was a direct ("no", "I don't know", head shaking) indication of no response. All sessions were tape-recorded and scoring was later verified from the recordings.

2.4. DATA ANALYSIS

Group data were analyzed with generalized linear mixed-effects modeling including crossed random effects for participants and verbs in a full random structure (Barr, Levy, Scheepers, & Tily, 2013), using a logit link function for binomial distributions (Dixon, 2008), employing function lmer of package lme4 (Bates, Maechler, & Bolker, 2011) in R (R Development Core Team, 2011).

Grammaticality judgment data were modeled as log odds of two types of responses ('resp': accept/reject) regressed onto two types of sentences ('sent': grammatical/ungrammatical) interacting with three functional category conditions ('cond': agreement/tense/aspect). In R notation this was specified as:

resp ~ sent*cond + (sent*cond|Person) + (sent*cond|Verb)

This approach obviates the need to calculate sensitivity indices (such as d′) because response probabilities are explicitly conditioned on sentence type so that hits are effectively compared to false alarms within the model.

Sentence production data were modeled as two types of outcomes (correct/incorrect) regressed onto three functional category conditions (agreement/tense/aspect). Productions were considered lexically correct when the appropriate verb stem (lexeme) was produced, regardless of conjugation. Following Varlokosta et al. (2006), productions were considered morphologically correct when the inflectional suffix was appropriate for the context with respect to the functional category being tested, regardless of other functional categories. For example, in sentences testing tense, responses were considered morphologically correct as long as the correct tense was produced regardless of agreement, aspect, or verb stem. Responses that were both lexically and morphologically correct, under these definitions, were considered overall correct.

3. Results

Figure 1 plots the error proportion by each group in both tasks on the top panels. Each box contains the middle half of the individual raw error counts (per person), while the whiskers extend to the entire range (performance of the persons with the least and greatest number of errors).

In grammaticality judgment, overall, patients made more errors in Tense and in Aspect than in Agreement (Tense: $\beta = 3.70$, $z = 3.34$, $p < .001$; Aspect: $\beta = 3.39$, $z = 4.08$, $p < .001$). There was no significant difference between Tense and Aspect ($\beta = 0.32$, $z = 0.48$, $p = .631$). However, not all errors are of the same kind: participants may accept an incorrect item or reject a correct item. Breaking down the errors along these lines (illustrated in Figure 1, middle and bottom panels), it turns out that the error difference arises entirely from acceptance of incorrect sentences (Tense vs. Agreement: $\beta = 3.44$, $z = 3.83$, $p < .001$; Aspect vs. Agreement: $\beta = 3.01$, $z = 3.25$, $p = .001$; Aspect vs. Tense: $\beta = 0.43$, $z = 0.73$, $p = .467$). There was no significant group difference in rejection of correct sentences (Tense vs. Agreement: $\beta = 0.31$, $z = 0.59$, $p = .554$; Aspect vs. Agreement: $\beta = 0.46$, $z = 1.18$, $p = .240$; Aspect vs. Tense: $\beta = 0.15$, $z = 0.37$, $p = .709$).

In sentence completion, there was no significant difference in patients' errors among any functional categories (Tense vs. Agreement: $\beta = 0.24$, $z = 0.69$, $p = .493$; Aspect vs. Agreement: $\beta = 0.39$, $z = 1.15$, $p = .249$; Aspect vs. Tense: $\beta = 0.15$, $z = 0.46$, $p = .648$). However, different kinds of errors are possible in this task as well: participants may produce an incorrect lexeme (a 'lexical error') or an incorrect inflection (a 'morphological error'). Breaking down the analysis accordingly, the lack of significant difference is replicated both for
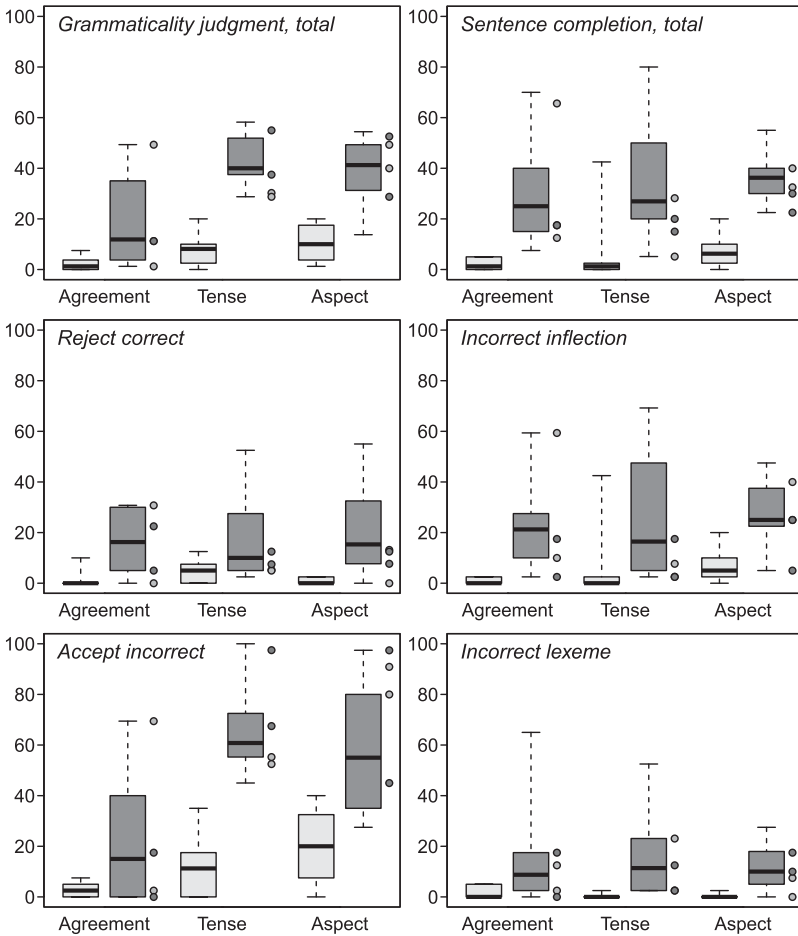
Fig. 1. Percentage of errors in each task and functional category condition by patients (dark bars) and control participants (light bars). Boxes denote interquartile range; thick lines mark the median; error bars extend to the full range. Filled circles next to the patient bar show the performance of patients with evidence for agrammatism (see text).

incorrect lexemes (Tense vs. Agreement: $\beta = 0.39$, $z = 1.13$, $p = .259$; Aspect vs. Agreement: $\beta = 0.28$, $z = 0.61$, $p = .543$; Aspect vs. Tense: $\beta = 0.11$, $z = 0.39$, $p = .700$) and for incorrect inflections (Tense vs. Agreement: $\beta = 0.05$, $z = 0.10$, $p = .920$; Aspect vs. Agreement: $\beta = 0.53$, $z = 1.47$, $p = .142$; Aspect vs. Tense: $\beta = 0.58$, $z = 1.35$, $p = .179$).

Therefore, the only significant differences among functional categories for patients arose from accepting incorrect Agreement less often than accepting incorrect Tense or incorrect Aspect. However, this difference was also observed

for the participants in the control group (Tense vs. Agreement: $\beta = 1.09$, $z = 2.24$, $p = .025$; Aspect vs. Agreement: $\beta = 1.92$, $z = 4.06$, $p < .001$), suggesting that it does not reflect a particular vulnerability or effect related specifically to aphasia. Additional significant differences for the control group were observed in acceptance of incorrect sentences (Tense vs. Aspect: $\beta = 0.84$, $z = 2.82$, $p = .005$) and in rejection of correct sentences in grammaticality judgment (Tense vs. Agreement: $\beta = 4.00$, $z = 2.31$, $p = .021$; Tense vs. Aspect: $\beta = 1.56$, $z = 2.28$, $p = .023$), as well as in incorrect inflections (Agreement vs. Aspect: $\beta = 3.00$, $z = 2.96$, $p = .003$; Tense vs. Aspect: $\beta = 1.87$, $z = 2.07$, $p = .038$) and incorrect lexemes (Agreement vs. Aspect: $\beta = 3.90$, $z = 2.54$, $p = .011$) in sentence completion.

A justified concern regarding group analyses is that they may conflate distinct underlying forms of impairment and thereby obscure individual patterns of particular theoretical importance. Turning to individual performance, Figure 2 plots the error proportions for both tasks on the top panels, for patients ranked in order of increasing severity (indexed by total number of errors in the indicated task). It is evident that there is no particular pattern of sentence completion performance with respect to the functional categories, at any severity region. In contrast, there is a relative lack of Agreement errors in grammaticality judgment, except for some of the most impaired patients. Focusing on the error types with the highest theoretical import, it is again evident that there is no systematic pattern of performance across functional categories, as seen in the bottom panels of Figure 2.

Table 3 lists the total number of errors for each patient in each task. Comparison of the individual relative proportions (via χ^2 tests, against homogeneous distribution) of error in the three functional categories indicated statistically significant lower proportions of grammaticality judgment errors in Agreement for the six less affected patients and the most affected one. Although additional sporadic significant differences were observed in both tasks, they would not survive a conservative adjustment criterion for multiple comparisons and they do not seem to pattern in any systematic way. Moreover, they are not consistent across comparably affected individuals, casting doubt on any potential individual interpretation.

The collected speech samples from the picture description task were analyzed following Faroqi-Shah and Thomson (2004). Specifically, we calculated mean length of phrase (MLP); proportion of grammatical phrases (out of the total number of phrases); and ratio of open-class to closed-class words. Following Varlokosta et al. (2006), a combination of reduced MLP and decreased proportion of grammatical phrases was considered evidence for agrammatism. Omission of function words, as evidenced by high ratio of open-class to closed-class words, is not considered a useful measure of agrammatism in a language with rich morphology, such as Greek.
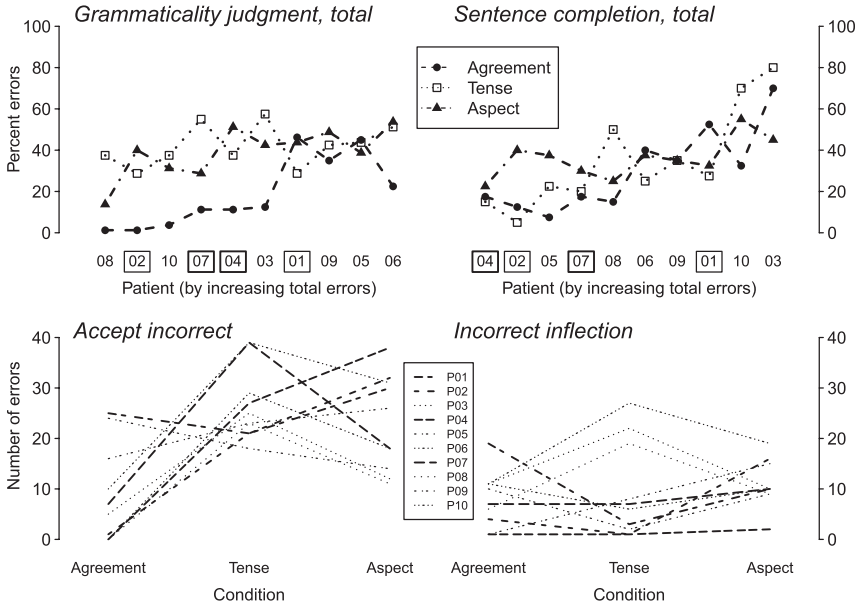
Fig. 2. Errors by each patient in each task and functional category condition. The same data points are plotted twice, to illustrate between-condition and between-patient variability. Top panels: percentage of errors by patient; lines join within-condition points. Bottom panels: number of errors (out of 240, in grammaticality judgment; out of 120, in sentence completion) by condition; lines join within-patient points. Patients with evidence for agrammatism (see text) are indicated with boxes in the top panels and bold lines in the bottom panels.

The comparison was made against the control group of Varlokosta et al. (2006), to maximize comparability with the previous findings, and also because picture description data from the present control group were not available. That control group did not differ from the present group of patients in age ($M = 57.3$, $SD = 11.2$, range 43–79; $t(11.88) = 0.878$, $p = .399$) or education ($M = 12.3$, $SD = 5.6$, range 3–18; $t(7.60) = 0.038$, $p = .971$). The results are shown in Table 4. Most patients presented with low proportions of grammatical phrases. Evidence for agrammatism, including low MLP, was strongest for Participants P04 and P07, followed by P02 and P01, who barely missed the $2SD$ cut-off for MLP (at -1.97 and -1.84, respectively). Participant P03, for whom picture description data were not available, exhibited low proportion of grammatical phrases (.69) in a narrative task (stroke story). As can be seen in the figures, the four patients with evidence for agrammatism are representative of the entire patient group, exhibiting individual variability spanning the group range, and do not constitute an identifiable subgroup in terms of the verb production and grammaticality judgment tasks.

TABLE 3. *Number of errors in each functional category for each participant, and test of homogeneous distribution of errors across categories*

Participant	Number of errors			Tns vs. Agr		Tns vs. Asp		Asp vs. Agr	
	Agr	Tns	Asp	χ^2	p	χ^2	p	χ^2	p
Grammaticality judgment[a]									
P01	37	23	35	3.27	.071	2.48	.115	0.06	.814
P02	1	23	32	20.17	< .001	1.47	.225	29.12	< .001
P03	10	46	34	23.14	< .001	1.80	.180	13.09	< .001
P04	9	30	41	11.31	.001	1.70	.192	20.48	< .001
P05	36	35	31	0.01	.906	0.24	.622	0.37	.541
P06	18	41	43	8.97	.003	0.05	.827	10.25	.001
P07	9	44	23	23.11	< .001	6.58	.010	6.12	.013
P08	1	30	11	27.13	< .001	8.80	.003	8.33	.004
P09	28	34	39	0.58	.446	0.34	.558	1.81	.179
P10	3	30	25	22.09	< .001	0.45	.500	17.29	< .001
Sentence completion[b]									
P01	21	11	13	3.12	.077	0.17	.683	1.88	.170
P02	5	2	16	1.29	.257	10.89	.001	5.76	.016
P03	28	32	18	0.27	.606	3.92	.048	2.17	.140
P04	7	6	9	0.08	.782	0.60	.439	0.25	.617
P05	3	9	15	3.00	.083	1.50	.221	8.00	.005
P06	16	10	15	1.38	.239	1.00	.317	0.03	.857
P07	7	8	12	0.07	.796	0.80	.371	1.32	.251
P08	6	20	10	7.54	.006	3.33	.068	1.00	.317
P09	14	14	14	0.00	1.000	0.00	1.000	0.00	1.000
P10	13	28	22	5.49	.019	0.72	.396	2.31	.128

NOTES: Agr = Agreeement; Tns = Tense; Asp = Aspect.
[a] out of 240 sentences (120 grammatical and 120 ungrammatical)
[b] out of 120 test sentences.

4. Discussion

In this study we have attempted to replicate the study of Varlokosta et al. (2006) regarding the functional categories of Greek verbs in aphasia. We have retained the task structure and design of the previous study but we have now controlled testing materials to the extent possible for amount of information (total, as well as preceding the verb), in an attempt to balance the processing requirements across conditions. We found little evidence for selective deficits: in sentence completion, performance of the participants with aphasia did not differ among the three conditions. In grammaticality judgment, performance in agreement was higher than in the other two functional categories, but this difference was also observed in the control group. Patients with the strongest evidence for agrammatism (based on a picture description task) did not cluster in any distinct way and did not individually exhibit any notable differences from the rest of the group. Therefore, processing considerations

TABLE 4. *Oral production measurements from a picture description task*

Participant	Total words	MLP	Proportion grammatical	Open:closed ratio
P01	73*	8.4	0.44*	0.97
P02	40*	8.0	0.60*	1.35*
P03[a]	–	–	–	–
P04	33*	5.5*	0.50*	1.06*
P05	160	22.9	0.71*	0.58
P06	117	13.0	1.00	0.89
P07	122	5.6*	0.36*	0.90
P08	110	15.6	0.57*	0.73
P09[a]	–	–	–	–
P10	37*	7.4*	1.00	0.68

NOTES: MLP = mean length of phrase.
[a] Data not available.
[*] more than 2 *SD* below the mean of the control group in Varlokosta et al. (2006).

not specific to aphasia apparently suffice to account for the current findings, as well as the previous findings of Varlokosta et al. Both sets of findings are thus inadequate for supporting structural approaches to agrammatism.

The pattern of performance observed in grammaticality judgment in both participant groups suggests that processing of tense and aspect is more demanding than processing of agreement in our testing materials. A simple possible explanation is apparent, taking into account that (a) agreement requires processing of a single word (i.e., the subject) and (b) this word immediately preceded the verb in all testing sentences. In contrast, temporal adverbs in the tense condition were sentence-initial, thus away from the verb, whereas the cues in the aspect condition were often multiword or far from the verb or both. Thus, superficial load considerations may suffice to account for the observed performance difference.

An alternative possible explanation based on linguistic constructs might take into account that agreement is a local relation in the sentence between the noun subject and the verb, whereas tense and aspect are interface categories presumably involving multiple processing of information at linguistic levels (Fyndanis et al., 2012; Varlokosta et al., 2006). In any case, a mild-to-moderate verbal processing deficit affecting processing span or capacity would not be expected to affect agreement performance. This is consistent with the observation that seven out of ten patients performed better in agreement than in the other two conditions. Apparently, deficits in aphasia accentuate a pre-existing difference in processing requirement among the categories. In other words, "what is linguistically complex is difficult for agrammatic speakers" (Bastiaanse et al., 2009, p. 25). Therefore, the relative 'impairment' in agreement cannot be characterized as being selective and due to agrammatism. A processing account of this sort, hinging

on linguistic constructs, could deflate the tension among 'processing' vs. 'structural' theorizing, acknowledging the need to accommodate both cognitive and linguistic theory in a comprehensive explanation of aphasia.

Median performance in sentence completion was lowest for aspect (see Figure 1, middle panel, right). This difference was statistically significant in the control group but not in the patient group. High individual variability among patients may have obscured a reliable underlying difference, which would be consistent with the aforementioned processing account. Correct production in the aspect condition requires coordination of multiple sources of information, including subject (for agreement), temporal adverb (for tense), and adverbial phrase (for aspect). Moreover, the aspect-specific indicator is often multiword and occasionally farther from the verb than the subject, inherently imposing maximal processing requirements. In this sense, a finding of impaired aphasic performance in the Aspect condition would hardly constitute evidence for a specific deficit. Therefore, it may be impossible to adequately control testing materials to allow unequivocal support of a structural/representational account in Greek.

The imbalance between functional category conditions in number of words intervening between the cue for the functional category and the verb is not easy to address, because more words are inherently needed to fully constrain the intended form in tense and particularly in aspect than in agreement. As Greek permits substantial flexibility in word order, a future study might manipulate the number of intervening words between critical constituents to examine the role of word distance in processing load without resulting in unusual sentence forms.

If impaired performance were due to a structural–representational deficit, it should manifest itself not only in grammaticality judgment but also in sentence completion for the same categories (Dickey, Milman, & Thompson, 2008). However, this was not the case in the present study, as the relatively higher performance in agreement was only observed in grammaticality judgment. Lower performance was generally observed in sentence completion than in grammaticality judgment across conditions, consistent with the higher processing demands of production, in which all features must be necessarily processed in order to achieve an acceptable response. Perhaps the lower processing demands of grammaticality judgment allowed better performance to surface in the relatively easier condition of Agreement. Alternatively, the explicit metalinguistic nature of grammaticality judgment may pose distinct constraints and stress different aspects of the processing system than production. If this is the case then it would not be appropriate to make a simple distinction between the two tasks simply as more or less demanding, and to expect greater differences or clearer deficits to emerge in production on the basis of this difference alone.

The findings of the present study point to a methodological issue regarding the reliability of testing individual patients to support or reject specific theoretical hypotheses. Individual cases are often considered most informative in neuropsychological and neurolinguistic research, for a number of well-founded reasons such as the possibility of very extensive testing, sample inhomogeneity plaguing group approaches, etc. (see discussion in Shallice & Buiatti, 2011, and references therein). Individual datapoints are considered stable and reliable indexes of performance, without regard to task or situation properties or to random fluctuations. However, the reliability of individual datapoints must be established prior to their employment in the evaluation of critical theoretical predictions; otherwise, non-replicable findings may misguide theoretical conclusions. This point can be illustrated by examining the performance of individual participants within the context of the entire group. For example, had we tested P08 only, we might have concluded in favor of a selective deficit primarily affecting tense.

Use of a control group is not informative in itself. Ceiling performance in a group of individuals without brain damage indicates nothing but lack of sensitivity of the task in detecting individual differences in this population. In a ceiling performance situation it cannot be determined whether in one condition top scores are barely achieved, perhaps with some effort, compared to other much easier conditions. For example, the agreement condition may be inherently easier than the aspect condition, in terms of linguistic complexity and associated requirements for cognitive resources, as argued above. However, if the task is overall too easy for the control group, so that sufficient cognitive resources are available for unimpaired participants to perform accurately in both conditions, then this important underlying difference will not emerge in the accuracy data. Lacking this distinction in the control data, one might misinterpret performance differences between conditions in the patient group as arising due to their aphasia, even though the differences pre-existed and were only exacerbated by a limitation on cognitive resources imposed by the disorder. To determine the relative difficulty of critical conditions, manipulations must be devised under which non-specific factors will bring control performance below ceiling, as minimum evidence of sensitivity. Only then can patterns of performance be meaningfully compared, to serve as baseline against which to consider selectivity of deficits in affected populations (cf. Dick, Bates, Wulfeck, Utman, Dronkers, & Gernsbacher, 2001).

Previous studies have not always controlled for overall sentence length or position of the verb within the sentence, and have not provided evidence for the equal difficulty of functional category processing by the control group. Friedmann and Grodzinky (1997) reported better performance in Agreement than in Tense in a single-case study in Hebrew, with materials controlled for

length but in the absence of control data. Wenzlaff and Clahsen (2004) found better performance in Agreement than in Tense in German, presenting control group data. However, their materials were not equated in phrase structure or length. Clahsen and Ali (2009) reported better performance in Agreement and Mood than in Tense in English, presenting control data, as well. However, sentences in the Mood condition were longer than in the Agreement and Tense conditions. Burchert, Swoboda-Moll, and De Bleser (2005) tested nine German agrammatic patients in Tense and Agreement, using materials controlled for length. They found no consistent pattern of performance and reported subgroups with different dissociations and levels of performance that cannot support general conclusions. Some patients performed at chance, for some patients Tense performance was higher than Agreement, while the opposite was the case for another patient. In Greek, Nanousi et al. (2006) found better performance in Agreement than in Tense, which in turn exhibited better performance than Aspect. Materials in the Tense and Agreement conditions were balanced but materials in the Aspect condition were longer. Thus, overall, and despite a rich literature investigating putative selective deficits in functional categories in agrammatism in a variety of languages, material construction has not always been sufficiently controlled and baseline performance is not established to be truly homogeneous.

Finally, it must be stressed that sentence length in itself is not an explanation for observed patterns of performance. That is, we are not offering a 'length account' for the observed performance differences among functional categories. We suggest that a specific processing account must be provided, in which processes will turn out to be affected by sentence length, to explain the observed differences. Other accounts, including structural ones, might also be compatible with the findings, as long as they can be sensitive to the sentence length manipulation. At the moment, length is merely a confound precluding a parsimonious explanation of theoretical import.

In sum, increased control of the informational load of the testing materials resulted in lack of differences among the three functional categories in sentence completion, and better performance in Agreement than in the other two categories in grammaticality judgment tasks. The performance of seven out of the ten persons with aphasia in grammaticality judgment tasks was consistent with the group results. Furthermore, the performance of the control participants as a group paralleled the performance of the participants with aphasia in grammaticality judgment tasks, consistent with an explanation based on the processing demands of the functional categories rather than a structural/representational deficit. It is concluded that further research is required into processing accounts of aphasia and that processing models must accompany any parsimonious account of aphasic performance.

REFERENCES

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, **68**, 255–278.

Bastiaanse, R., Bouma, G., & Post, W. (2009). Linguistic complexity and frequency in agrammatic speech production. *Brain & Language*, **109**, 18–28.

Bastiaanse, R., & Grodzinsky, Y. (2000). *Grammatical disorders in aphasia: a neurolinguistic perspective*. London: Whurr.

Bastiaanse, R., & Thompson, C. K. (2012). *Perspectives on agrammatism*. New York: Psychology Press.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: linear mixed-effects models using S4 classes. R package version 0.999375-42, online <http://CRAN.R-project.org/package=lme4>.

Beretta, A. (2008). Disorders of syntax. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 155–163). New York: Elsevier.

Burchert, F., Swoboda-Moll, M., & De Bleser, R. (2005). Tense and Agreement dissociations in German agrammatic speakers: Underspecification vs. Hierarchy. *Brain and Language*, **94**, 188–199.

Clahsen, H., & Ali, M. (2009). Formal features in aphasia: Tense, Agreement, and Mood in English agrammatism. *Journal of Neurolinguistics*, **22**, 436–450.

Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., & Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review*, **108**, 759–788.

Dickey, M. W., Milman, L. H., & Thompson, C. K. (2008). Judgment of functional morphology in agrammatic Broca's aphasia. *Journal of Neurolinguistics*, **21**, 35–65.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, **59**, 447–456.

Faroqi-Shah, Y., & Thompson, C. K. (2004). Semantic, lexical, and phonological influences on the production of verb inflections in agrammatic aphasia. *Brain and Language*, **89**, 484–498.

Friedmann, N., & Grodzinsky, Y. (1997). Tense and agreement in agrammatic production: pruning the syntactic tree. *Brain & Language*, **56**, 71–90.

Fyndanis, V., Varlokosta, S., & Tsapkini, K. (2012). Agrammatic production: interpretable features and selective impairment in verb inflection. *Lingua*, **122**, 1134–1147.

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

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The assessment of aphasia and related disorders* (3rd ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.

Holton, D., Mackridge, P., & Philippaki-Warburton, I. (1997). *Greek: a comprehensive grammar of the modern language*. London: Routledge.

Kean, M. L. (1977). The linguistic interpretation of aphasic syndromes: agrammatism in Broca's aphasia, an example. *Cognition*, **5**, 9–46.

Kean, M. L. (Ed.) (1979). Agrammatism: A phonological deficit? *Cognition*, **7**, 69–84.

Kok, P., van Doorn, A., & Kolk, H. (2007). Inflection and computational load in agrammatic speech. *Brain and Language*, **102**, 273–283.

Menn, L., & Obler, L. (1989). *Agrammatic aphasia: a cross-language narrative sourcebook*. Philadelphia: John Benjamins.

Nanousi, V., Masterson, J., Druks, J. & Atkinson, M. (2006). Interpretable vs. uninterpretable features: evidence from six Greek-speaking agrammatic patients. *Journal of Neurolinguistics*, **19**, 209–238.

R Development Core Team (2011). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, online <http://www.R-project.org/>.

Ralli, A. (1988). *Eléments de la Morphologie du Grec Moderne: La Structure de la Verbe*. Unpublished PhD dissertation, University of Montreal.

Shallice, T., & Buiatti, T. (2011). Types of case series–the anatomically based approach: commentary on M. F. Shwartz & G. S. Dell: Case series investigations in cognitive neuropsychology. *Cognitive Neuropsychology*, **28**, 500–514.

Tsapkini, K., Jarema, G., & Kehayia, E. (2002). Regularity revisited: evidence from lexical access of verbs and nouns in Greek. *Brain and Language*, **81**, 103–119.

Tsapkini, K., Vlahou, C. H., & Potagas, C. (2009/2010). Adaptation and validation of standardized aphasia tests in different languages: lessons from the Boston Diagnostic Aphasia Examination–Short Form in Greek. *Behavioural Neurology*, **22**, 111–119.

Varlokosta, S., Valeonti, N., Kakavoulia, M., Lazaridou, M., Economou, A., & Protopapas, A. (2006). The breakdown of functional categories in Greek aphasia: evidence from agreement, tense, and aspect. *Aphasiology*, **20**, 723–743.

Wenzlaff, M., & Clahsen, H. (2004). Tense and agreement in German agrammatism. *Brain & Language*, **89**, 57–68.