Grammaticality Judgment in Aphasia: Deficits Are Not Specific to Syntactic Structures, Aphasic Syndromes, or Lesion Sites

Stephen M. Wilson[1] and Ayşe Pınar Saygın[2]

Abstract

■ We examined the abilities of aphasic patients to make grammaticality judgments on English sentences instantiating a variety of syntactic structures. Previous studies employing this metalinguistic task have suggested that aphasic patients typically perform better on grammaticality judgment tasks than they do on sentence comprehension tasks, a finding that has informed the current view that grammatical knowledge is relatively preserved in agrammatic aphasia. However, not all syntactic structures are judged equally accurately, and several researchers have attempted to provide explanatory principles to predict which structures will pose problems to agrammatic patients. One such proposal is Grodzinsky and Finkel's (1998) claim that agrammatic aphasics are selectively impaired in their ability to process structures involving traces of maximal projections. In this study, we tested this claim by presenting patients with sentences with or without such traces, but also varying the level of difficulty of both kinds of structures, assessed with reference to the performance of age-matched and young controls. We found no evidence that agrammatic aphasics, or any other subgroup, are selectively impaired on structures involving traces: Some judgments involving traces were made quite accurately, whereas other judgments not involving traces were made very poorly. Subgroup analyses revealed that patient groups and age-matched controls had remarkably similar profiles of performance across sentence types, regardless of whether the patients were grouped based on Western Aphasia Battery classification, an independent screening test for agrammatic comprehension, or lesion site. This implies that the pattern of performance across sentence types does not result from any particular component of the grammar, or any particular brain region, being selectively compromised. Lesion analysis revealed that posterior temporal areas were more reliably implicated in poor grammaticality judgment performance than anterior areas, but poor performance was also observed with some anterior lesions, suggesting that areas important for syntactic processing are distributed throughout the left peri-sylvian region. ■

INTRODUCTION

Much of the recent literature on syntactic comprehension in aphasia focuses on Broca's aphasia. This was originally believed to be primarily an expressive disorder, since these patients' comprehension appears to be relatively intact in comparison with their effortful, disorganized speech (Goodglass, 1993). However, a number of important studies in the 1970s and 1980s demonstrated that when all semantic and pragmatic cues to meaning are eliminated so that comprehension must rely on syntactic structure alone, Broca's aphasics perform very poorly. For instance, Caramazza and Zurif (1976) showed that Broca's aphasics experience difficulty with center-embedded object relative clauses when forced to rely on syntax alone, and Schwartz, Saffran, and Marin (1980) found that a group of Broca's aphasics performed at chance on reversible passive structures. Findings such as these led to the claim that Broca's aphasics are "agrammatic," and it was widely held that their expressive and comprehension deficits stemmed from a common source: an impaired ability to process grammatical structure.

This interpretation was challenged in a seminal study of grammaticality judgment in agrammatic aphasia carried out by Linebarger, Schwartz, and Saffran (1983). In this study, four patients who performed very poorly on comprehension tasks relying on syntax were found to have excellent performance in a grammaticality judgment task. Linebarger et al. presented compelling arguments that success in their task could not have been achieved unless the patients in fact had considerable syntactic knowledge still intact. One possible explanation suggested by Linebarger et al. for the discrepancy between performance on comprehension and grammaticality judgment tasks is that these patients can compute syntactic structures, but they cannot map them onto a semantic interpretation. In any case, the finding that Broca's aphasics perform quite well on grammaticality judgment tasks has been replicated in a number of studies involving a wide variety of constructions (e.g., Wulfeck & Bates, 1990; Wulfeck, Bates, & Capasso, 1991; Schwartz,

[1]University of California, Los Angeles, [2]University of California, San Diego

Linebarger, Saffran, & Pate, 1987) and in a range of languages, such as Serbo-Croatian (Lukatela, Crain, & Shankweiler, 1988), Italian (Wulfeck et al., 1991), and Mandarin Chinese (Lu et al., 2000).

However, another consistent finding that has emerged from this literature is that while grammaticality judgment is relatively intact, not all structures are judged equally well. For instance, Linebarger et al. (1983) found that patients were relatively poor at detecting violations of gender agreement (1a) and auxiliary agreement (1b) in tag questions. Note that according to linguistic convention, ungrammatical sentences are preceded by an asterisk.

(1a) *The little boy fell down, didn't it?
(1b) *John is very tall, doesn't he?

Wulfeck et al. (1991) systematically compared English- and Italian-speaking agrammatic Broca's aphasics' sensitivity to errors of agreement and word order, and found that patients were poorer at detecting agreement errors in both languages (although Italian patients were better at detecting agreement errors than their English counterparts, in line with crosslinguistic differences also observed in normal subjects). Such patterns can be interpreted in the light of wider theories of language breakdown. For instance, Wulfeck et al. argued that the relative difficulty patients had in detecting agreement errors reflects the fact that morphological marking is a ''weak link'' in language processing in general, thus it is particularly vulnerable in aphasia as well as for normal subjects under nonoptimal conditions such as increased processing load or acoustic degradation (Dick et al., 2001; Blackwell & Bates, 1995).

Some researchers have proposed that the locus of impairment in agrammatism is very specific, affecting only a single component of the grammar (e.g., Mauner, Fromkin, & Cornell, 1993; Grodzinsky, 1986, 2000; Grodzinsky & Finkel, 1998). In this article, we will focus in particular on the claim by Grodzinsky and Finkel (1998) that agrammatic aphasics are selectively impaired in their ability to process structures involving traces of maximal projections (defined below), therefore, they will perform poorly on grammaticality judgments that depend crucially on attention to such traces. Grodzinsky and Finkel (1998) present experimental results that they claim support their proposal, and their findings have been interpreted by linguists, such as Chomsky (2000), as providing evidence for a neural distinction corresponding to the linguistic distinction made in several current syntactic theories between movement (defined below) of different kinds of elements.

Before continuing, we will outline some basic theoretical concepts assumed by researchers, such as Grodzinsky and Finkel (1998), working within the GB/minimalist generative linguistic tradition (Chomsky, 1995). It has long been observed that certain linguistic structures are in some sense related to one another, such as the active (2a) and passive (2b) sentences below:

(2a) The dog chased the cat.
(2b) The cat was chased by the dog.

In particular, although the surface word orders differ, and the sentences differ in terms of information structure and pragmatics, they are identical in terms of thematic roles, or ''who did what to whom.'' To capture this similarity, most current generative theories assume that there is some level of structure at which these two sentences are identical or at least similar, a level that encodes the thematic roles of the participants. In GB/minimalism, the noun phrase *the cat* is assumed to follow the verb *chase* at this underlying level of structure. This order is preserved in the surface structure in (2a), whereas in (2b), *the cat* is moved in front of the verb. However, when *the cat* is moved, it leaves a *trace* in the position where it originated. The moved noun phrase *the cat* and its trace together constitute a *chain*, and constraints on movement are generally stated as constraints governing the validity of chains. Traces play in important role in GB/minimalism: They are considered to be psychologically real units just as words and morphemes are, although they do not have any overt phonetic realization. Thus the structure of (2b) is assumed to be something like:

(3) The cat$_i$ was chased t_i by the dog.

The *i* subscripts relate the trace *t* to *the cat*. In this case, the element that has moved is a complete noun phrase. Complete phrases, such as this complete noun phrase, are also referred to as *maximal projections*. It is the traces of maximal projections specifically that Grodzinsky and Finkel (1998) claim to be selectively impaired in agrammatic aphasia. This is referred to as the *trace deletion hypothesis* (TDH). Numerous related claims have been made in the recent literature (e.g., Beretta, Pinango, Patterson, & Harford, 1999; Mauner et al., 1993; Grodzinsky, 1986, 2000). We will refer to traces of maximal projections as *XP traces*, where *X* stands for any category (e.g., noun, verb) and *P* stands for phrase.

It is also possible for elements other than maximal projections to undergo movement. Consider the following sentences:

(4a) Mike was a good student.
(4b) Was Mike a good student?

In this case, the element that has moved is *was*, which is an copula verb. It is not a whole phrase itself, but rather it is the head of the verb phrase *was a good student*. Thus the movement in (4b) is termed *head movement*, and the trace left behind by *was* is not a trace of a maximal projection (an XP trace), but rather a trace of a head. According to the TDH, structures such as (4b) should pose no difficulty for agrammatic aphasics, since they fall outside the hypothesized locus of impairment.

The experimental evidence Grodzinsky and Finkel (1998) bring to bear on this hypothesis consists of grammaticality judgments by Broca's and Wernicke's aphasics on sentences involving eight different kinds of violations. Four of these are claimed to involve XP traces, whereas the other four are claimed to be violations of other grammatical principles assumed by the TDH to be intact. For each violation type, grammatical and ungrammatical sentences were presented, and performance was assessed as the sum of performance on grammatical and ungrammatical sentences, that is, the combined ability to correctly accept the former and correctly reject the latter. Examples of two contrasts held by Grodzinsky and Finkel (1998) to depend upon XP traces are shown in (5) and (6):

(5a) John seems likely to win.
(5b) *John seems that it is likely to win.
(6a) Which woman did David think saw John?
(6b) *Which woman did David think that saw John?

Example (5) shows a violation of a condition on chains called *relativized minimality* (Rizzi, 1990), which prohibits any element of like kind intervening between a trace and its antecedent. In this case, the pronoun *it* intervenes between *John* and the trace of *John*, which would be in front of the verb *win*. Because the chain involves a trace of a maximal projection, judgment of the violation is argued to be impossible for agrammatic aphasics. The contrast in (6) is an example of a pattern called the *that-trace effect*. In GB/minimalism, it is assumed that the ungrammaticality of (6b) results from an interaction between the complementizer *that* and the trace of *which woman* (a maximal projection), which is assumed to move through a position immediately to the left of *that* on its way to the front of the sentence. It is beyond the scope of this article to fully explicate the theoretical arguments which motivate these analyses. The important point is that the ungrammaticality of (5b) and (6b) is thought to result from violations of principles involving XP traces.

Examples (7) and (8) show violations of grammatical principles that arguably do not involve traces of maximal projections.

(7a) The children threw the football over the fence.
(7b) *The children sang the football over the fence.
(8a) Could they have left town?
(8b) *Have they could left town?

Example (7b) is ungrammatical because the lexical requirements placed by the verb *sing* on its arguments are violated, that is, that the direct object, if present, should be the song sung. A football is not a song, so it cannot be sung. Example (8) is another violation of relativized minimality, but involving a trace of a head rather than a trace of a maximal projection. The declarative sentence from which this interrogative is derived would be:

(8c) They could have left town.

The grammatical question in (8a) is formed by moving the closest auxiliary verb *could* to the front of the sentence. (8b) is held to be ungrammatical because the auxiliary *have* crosses another auxiliary verb, namely *could*, thus violating relativized minimality.

Having argued that the contrasts in (5) and (6) depend crucially on XP traces, whereas the contrasts in (7) and (8) do not, Grodzinsky and Finkel (1998) evaluated the performance of aphasic patients on grammaticality judgments of these sentences, along with two other types of violations in each condition. They found that both Broca's and Wernicke's aphasics performed considerably worse on the sentences involving traces of maximal projections. Specifically, Broca's aphasics (n = 4) had an error rate of 40% on the sentences involving XP traces compared to 13% on the other sentences, and the error rates for Wernicke's aphasics (n = 7) were 34% versus 19%. The main effect of sentence type was significant, the main effect of patient group was not significant, and the interaction of sentence type by patient group approached significance (p = .078). Grodzinsky and Finkel (1998) draw several conclusions from these findings. The first, which appears to be indisputable, is that both Broca's and Wernicke's aphasics do have substantial deficits in grammaticality judgment, in contrast to earlier findings, such as those of Linebarger et al. (1983). Second, they claim that these deficits are restricted to structures involving traces of maximal projections, whereas performance on other structures is intact, in line with the predictions of the TDH. Third, they claim a between-groups difference, although as noted above, the interaction did not in fact reach significance.

There are two substantial methodological problems with this study that cast doubt on the conclusion that the disruption is structure-specific. First, the sentences in the XP trace condition differ in many ways from those in the "other" condition, and pretheoretically, the judgments in the XP trace condition simply appear to be much subtler and difficult than those in the "other" condition. In fact, for one of the XP trace contrasts ("superiority"), we do not even share Grodzinsky and Finkel's (1998) grammaticality judgments:

(9a) When did John do what?
(9b) *What did John do when?

Contrary to their assessment of (9b) as ungrammatical, we find both sentences equally acceptable. Looking ahead to the results of our study, we can report that normal subjects usually either accept both of these sentence types or reject both, rather than showing the differentiation claimed by Grodzinsky and Finkel (1998).

This leads to the second significant methodological shortcoming: lack of appropriate control subjects. Although Grodzinsky and Finkel (1998) report informally that they tested "several neurologically intact, socioeconomically and age-matched control subjects, whose error rates were less than 5%, randomly distributed across conditions" (p. 310), they do not provide any details on this process, and whatever control testing they did carry out was clearly insufficient, as the judgment problem just noted attests to, and the results for control subjects in the present study will demonstrate further.

To more appropriately assess aphasic patients' grammaticality judgment performance on various syntactic structures, we carried out a study similar to Grodzinsky and Finkel's (1998), but with two crucial modifications. First, in addition to aphasic patients, we tested college-aged and age-matched control subjects with the same experimental design. Besides, Broca's and Wernicke's aphasics, we also tested a group of anomic aphasics, who show much milder deficits, to further explore the relationship between performance profile and aphasia subtype.

Second, we added two conditions: sentences where the grammaticality judgment crucially depends upon an XP trace, yet the contrasts are pretheoretically "easy," and conversely, sentences with other types of violations that were intuitively more subtle and "hard." These assessments of easiness versus difficulty were initially made based on our own intuitions in designing the stimuli, however they are borne out by the performance both of control subjects and of patients.

We tested two contrasts that we argue depend crucially on XP traces. These are shown in (10) and (11).

(10a) The dog which bit me was black.
(10b) *Me the dog which bit was black.
(11a) What did Bill buy besides apples?
(11b) *What did Bill buy apples and?

Although the sentences in (10b) and (11b) are very obviously ungrammatical, we argue that within the GB/minimalist framework, they involve solely violations of principles relating to traces of maximal projections, so according to the TDH, judging these sentences should be difficult for agrammatic aphasics. Example (10b) is formed from (10a) via a process called *topicalization*. Note that in general, topicalization is a possible operation in English:

(12) I hate apples, but bananas I like!

In this example, *bananas*, which is the object of *like*, has been topicalized and moved in front of the subject of that clause (*I*). The reason why topicalization is impossible in (10b) is that *me* starts out inside a relative clause (*which bit me*). It is impossible to move anything out of a relative clause, and in GB/minimalism, this traditionally follows from a constraint called *subjacency*. Basically, subjacency forces movement to take place step-by-step, and *me* in (10) cannot move step-by-step to the front of the clause since the word *which* blocks one of the crucial intermediate landing sites. Hence, the subjacency violation cannot be detected without access to the trace of *me*, because the violation results from the inability of this trace and the complementizer *which* to occupy the same position.

Example (11b) is ungrammatical because it violates the *coordinate structure constraint*, which states roughly that one member of a pair of conjuncts cannot move. In terms of chains and traces, the condition would be stated as ruling out chains where a trace appears in a conjunct, so again detection of the violation requires access to XP traces.

We also tested two contrasts that are intuitively fairly subtle yet which do not rely on XP traces:

(13a) Could they have left without me?
(13b) *Could have they left without me?
(14a) She donated the books to the library.
(14b) *She donated the library the books.

Example (13b) could be analyzed in several ways, but there is no doubt that it involves head movement and therefore traces of heads, which under the TDH are assumed to be intact in agrammatic aphasia, as explained above. Example (14b) contains an argument structure violation similar to (7). However, (14b) and other similar examples we used are made more difficult by the fact that they are very similar to grammatically well-formed structures. In this case, the ditransitive argument structure is permitted by the semantically related verb *give*:

(15) She gave the library the books.

In summary then, we tested four conditions, which are outlined and exemplified in Table 1, along with the condition names we will use for convenience. The sentences in the trace/hard and other/easy conditions were based on those used by Grodzinsky and Finkel (1998). Patients and control subjects performed grammaticality judgments on these structures. College-aged and age-matched control groups were also tested on the other four sentence types from Grodzinsky and Finkel (1998), although to limit the duration of the experiment, we did not test patients on these. Note that in our terms, Grodzinsky and Finkel (1998) tested only the trace/hard and other/easy conditions; by including a wider range of structures, we aimed to avoid confounding dependence on XP traces with intrinsic difficulty, and by testing control subjects with the same experimental design as patients, we could determine baseline performance on each sentence type. Finally, a lesion analysis was carried out to identify areas where damage

Table 1. Conditions and Example Sentences for Each Condition

Condition	*Set*	*Grammatical*	*Ungrammatical*
Trace/Hard	1	David seems likely to win.	*John seems that it is likely to win.
	2	Which woman did David think saw Pete?	*Which woman did John think that saw Tony?
Trace/Easy	1	The dog which bit me was black.	*Me the dog which bit was black.
	2	What did Bill buy besides apples?	*What did Bill buy oranges and?
Other/Hard	1	Could they have left without me?	*Could have they left without us?
	2	He donated the books to the library.	*She donated the library the books.
Other/Easy	1	The children threw the football over the fence.	*The children sang the football over the fence.
	2	Could they have left town?	*Have they could left the city?

was associated with deficits in various aspects of task performance.

RESULTS

Controls

The results for college-age controls (n = 26) and age-matched controls (n = 14) in terms of percent correct summed across grammatical and ungrammatical structures are shown in Figure 1. These results include only the sentence types on which the aphasic patients were also tested; the additional sentence types used by Grodzinsky and Finkel (1998) are reported separately at the end of this section.

A one-between (control group) by one-within (sentence type) repeated measures ANOVA revealed a main effect of sentence type, $F(3,114) = 21.8$, $p < .0001$, but no main effect of group, $F(1,38) = 2.08$, $p = .16$, nor a significant interaction of group by sentence type, $F(3,114) = 2.20$, $p = .11$. The order of difficulty among the sentence types was trace / hard (least squares mean = 84%) > other / hard (88%) > other/easy (93%) > trace/easy (95%). Planned comparisons showed that performance was better on trace/easy than trace/hard, and better on other/easy than other/hard, that trace/hard was harder than other/hard, and that trace/easy was easier than other/easy (Bonferroni-corrected linear contrasts with pooled variance, all ps $\leq$ .05). There was a tendency for the age-matched group to perform more poorly, most notably on the trace/hard condition, on which their performance was just 80%.

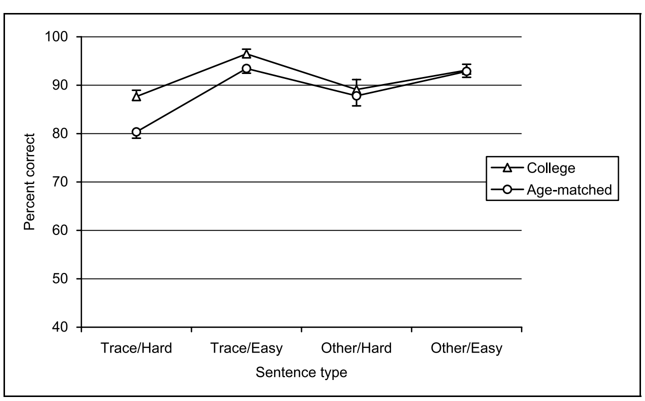

Figure 1. Grammaticality judgment performance of college-aged and age-matched control subjects across the four sentence types. Error bars on this and all subsequent figures are plus or minus one standard error of the mean.

When we included four additional sentence types from Grodzinsky and Finkel (1998) (which we did not test on aphasic patients), the least squares mean for the trace/hard condition decreased substantially from 84% to 78%, while the least squares mean for the other/easy condition increased slightly from 93% to 95%. The drops in trace/hard performance were driven by performance on the ''superiority'' contrasts exemplified in (9). Both control groups achieved just 60% correct on these sentences, although note that we here we defined ''correct'' as ''in accordance with Grodzinsky and Finkel's (1998) judgments,'' which as noted above we do not share.

In sum, for control subjects, there were highly significant differences between sentence types. The large differences found in normal subjects' performance on the trace/hard versus other/easy conditions fail to replicate Grodzinsky and Finkel's (1998) claim that normal performance on these sentences in uniformly above 95%, and indicate that their experimental conditions were not matched for level of difficulty. The college-aged controls were tested to determine ''optimal'' performance levels, but it would not be appropriate to compare their performance directly to that of the aphasic patients. However, the results for age-matched controls provide baseline measures against which we can evaluate the performance of aphasic patients in subsequent sections.

Aphasic Patients Grouped by Western Aphasia Battery

Information about the aphasic patients is summarized in Table 2, along with their raw scores on the four

Table 2. Characteristics of Patients Who Participated in the Study, Along With Their Raw Scores

Patient	*Age*	*Hnd*	*WAB*	*Syntactic comprehension*	*pIFG lesioned*	*pTemp lesioned*	*T/H (%)*	*T/E (%)*	*O/H (%)*	*O/E (%)*
B. K.	57	R	Anomic	Agrammatic	No	No	50	75	50	71
C. H.	67	R	Anomic	Good	No	No	63	96	67	79
D. D.	57	R	Broca's	Poor	Yes	Yes	50	83	71	75
F. Y.	79	R	Wernicke's	Poor	No	Yes	67	79	58	75
H. K.	64	R	Wernicke's	Good	Yes	No	50	63	54	58
H. M.	74	R	Broca's	Poor	Yes	Yes	46	46	54	38
J. B.	67	R	Broca's	Agrammatic	Yes	–	46	58	67	63
J. C.	81	R	Anomic	Good	–	–	71	92	71	88
J. H.	64	L	Anomic	Good	Yes	No	83	100	88	96
J. S.	52	R	Broca's	Agrammatic	Yes	–	63	79	63	92
J. T.	78	L	Wernicke's	Agrammatic	No	Yes	46	63	67	63
J. W.	74	R	Anomic	Poor	No	Yes	58	96	54	83
K. W.	66	R	Anomic	Good	Yes	No	67	92	71	75
L. R.	57	R	Anomic	Poor	Yes	Yes	54	79	58	75
M. B.	51	R	Broca's	Poor	No	No	46	79	58	71
P. B.	76	R	Anomic	Good	No	No	92	96	88	96
P. P.	51	L	Wernicke's	Poor	Yes	Yes	58	50	63	58
R. S.	76	R	Wernicke's	Poor	No	Yes	46	63	38	58
V. H.	73	L	Wernicke's	Good	Yes	–	54	75	54	71
W. G.	83	R	Wernicke's	Agrammatic	No	Yes	54	58	33	46
W. R.	59	R	Broca's	Poor	Yes	No	67	83	67	88
W. T.	67	R	Wernicke's	Poor	Yes	Yes	46	83	54	63
Coll.	18–36	N/A	N/A	N/A	N/A	N/A	88 ± 1	96 ± 1	89 ± 2	93 ± 1
A. M.	54–79	N/A	N/A	N/A	N/A	N/A	80 ± 2	93 ± 2	88 ± 4	93 ± 2

Abbreviations: Hnd = handedness; WAB = classification by the Western Aphasia Battery; pIFG = posterior inferior frontal gyrus; pTemp = posterior temporal region; T/H = average score on trace/hard condition; T/E = trace/easy; O/H = other/hard; O/E = other/easy; Coll. = college-aged controls; A.M. = age-matched controls. Standard errors of the mean are shown for control groups' scores.

conditions. The performance of patients grouped by the Western Aphasia Battery (WAB; Kertesz, 1982) into Broca's (n = 6), Wernicke's (n = 8), and anomic (n = 8) groups is depicted in Figure 2, along with age-matched controls (n = 14) who provide a reference point for this and several subsequent comparisons. A one-within (group) by one-between (sentence type) repeated measures ANOVA revealed a main effect of group, $F(3,32) = 21.6$, $p < .0001$, with all three aphasic groups performing worse than age-matched controls (Tukey's HSD). Among the aphasic groups, the only comparison that was significant was between anomics, the least impaired group, and Wernicke's aphasics, the most impaired group. There was a main effect of sentence type, $F(3,96) = 43.3$, $p < .0001$, with the order of difficulty among the sentences being trace/hard (least squares mean = 63%) > other/hard (68%) > other/easy (77%) > trace/easy (81%). Planned comparisons produced the same results as with the control groups: performance was better on trace/easy than trace/hard, and better on other/easy than other/hard; trace/hard was harder than other/hard; and trace/easy was easier than other/easy (Bonferroni-corrected linear contrasts with pooled variance, all ps $\leq .05$).

The interaction of group by sentence type was also significant, $F(9,96) = 2.24$, $p = .034$, indicating that the pattern of impairment across groups was not identical. To explore this interaction further, we carried out three sub-ANOVAs comparing each of the patient groups to the age-matched control group. For the Broca's aphasics, there was a main effect of group,

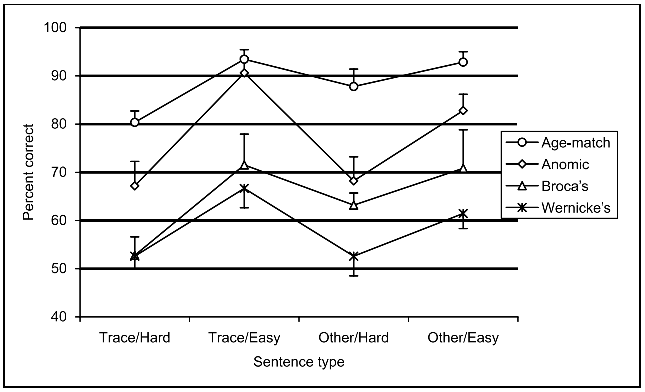

Figure 2. Grammaticality judgment performance of patients classified by the WAB into anomic, Broca's, and Wernicke's groups and age-matched control subjects across the four sentence types.

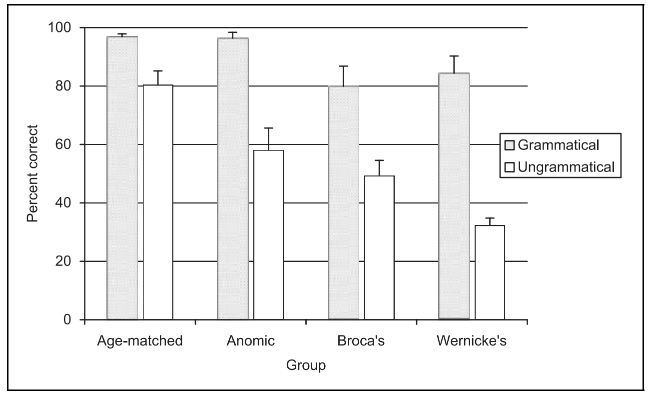

Figure 3. Grammaticality judgment performance of patients classified by the WAB and age-matched control subjects on grammatical versus ungrammatical sentences.

$F(1,18) = 28.3$, $p < .0001$, and a main effect of sentence type, $F(3,54) = 19.4$, $p < .0001$, but no interaction of group by sentence type, $F(3,54) = 0.632$, $p = .58$. The results for the Wernicke's aphasics were similar, with a main effect of group, $F(1,20) = 75.2$, $p < .0001$, and a main effect of sentence type, $F(3,60) = 17.2$, $p < .0001$, but again no interaction, $F(3,60) = 1.62$, $p = .20$. For the anomic patients, there was a main effect of group, $F(1,20) = 7.59$, $p = .012$, a main effect of sentence type, $F(3,60) = 43.8$, $p < .0001$, and a significant interaction of group by sentence type, $F(3,60) = 7.39$, $p = .0004$. As can be seen in Figure 2, anomic patients were relatively more impaired on the other/hard sentences, and relatively less impaired on the trace/easy sentences. We are reluctant to over-interpret this pattern as it does not follow from the predictions of any theory of which we are aware.

These results are surprising, since they suggest that neither Broca's aphasics nor Wernicke's aphasics are differentially impaired on any particular sentence type in grammaticality judgment. This argues against the claims of Grodzinsky and Finkel (1998) that these groups are selectively impaired on structures involving XP traces, and also against the possibility that patients could be relatively more impaired on judging the grammaticality of the more difficult structures, as appears to be the case in sentence comprehension tasks (Dick et al., 2001).

Most of the patients' errors were incorrect acceptances of ungrammatical sentences. This can be seen in Figure 3, which depicts the performance of each group on grammatical and ungrammatical sentences.

Aphasic Patients Grouped by Grammatical Comprehension Performance

In this section, instead of grouping patients based on the WAB classification, we grouped them based on their performance on a grammatical comprehension task involving active and passive sentences. Subjects heard reversible sentences, such as "The cat is biting the dog" or "The dog is bitten by the cat" and were asked to choose a picture corresponding to the animal that was "doing the bad action" in each sentence. Performance on each of the two conditions (active, passive) was characterized as being at chance, above chance, or below chance (binomial test, 95% confidence interval, two-tailed). We defined *agrammatic* according to a common definition in the literature (e.g., Grodzinsky, 2000), whereby performance must be above chance for active sentences, but at chance for passive sentences. There were five patients who met these criteria. The remaining patients were divided into "good comprehenders" ($n = 7$), who were above chance in both conditions, and "poor comprehenders" ($n = 10$), who were either at chance or below in both conditions, or who achieved one of the other logically possible outcomes, which sometimes arise probably when patients employ nonsyntactic strategies. The results for these three groups, compared to age-matched controls ($n = 14$), are shown in Figure 4.

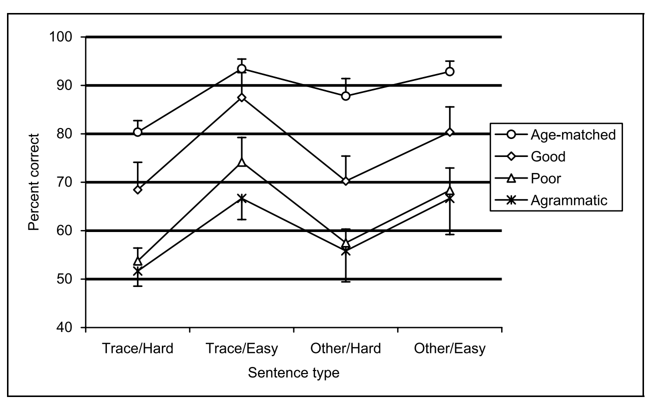

Figure 4. Grammaticality judgment performance of patients classified by an independent screening test for agrammatic comprehension and age-matched control subjects across the four sentence types.

A repeated measures ANOVA revealed a main effect of comprehension-defined group, $F(3,32) = 16.9$, $p < .0001$, with age-matched controls performing the best (mean = 89%), followed by good comprehenders (77%), poor comprehenders (63%), and agrammatics (60%). Pairwise comparisons were significant between controls versus poor comprehenders, controls versus agrammatics, and good comprehenders versus agrammatics (Tukey's HSD). There was a main effect of sentence type, $F(3,96) = 36.4$, $p < .0001$. However, there was no interaction of comprehension group by sentence type, $F(9,96) = 1.14$, $p = .35$. This implies that aphasic patients who are classified as agrammatic by the performance on an independent comprehension test do not show a pattern of performance on this grammaticality judgment task, which differs from any other group. Rather, they are more impaired across the board on all sentence types.

It has been suggested that Broca's aphasics with agrammatic comprehension defined as above constitute a special subgroup (Grodzinsky, Pinango, Zurif, & Drai, 1999). In our study, two of the six Broca's aphasics evidenced agrammatic comprehension (J.B. and J.S.). We further examined the grammaticality judgment performance of the two agrammatic Broca's aphasics. J.B. scored 46% on the trace/hard condition, 58% on trace/easy, 67% on other/hard, and 63% on other/easy. Although he tended to do worse on sentences involving XP traces, this was not significant (Fisher's Exact Test, two-tailed, $p = .30$). In fact, his performance was not significantly different from chance on any of the four conditions, assessed by comparing his proportion of ''accept'' responses across grammatical versus ungrammatical sentences (Fisher's Exact Test, one-tailed, all ps $> .05$). The second patient, J.S., scored 63% on the trace/hard condition, 79% on trace/easy, 63% on other/hard, and 92% on other/easy. Her totals on the trace conditions are marginally lower than on the other conditions, but this difference was not significant (Fisher's Exact Test, two-tailed, $p = .64$). Her performance on both of the easy conditions was above chance (calculated as above, Fisher's Exact Test, one-tailed, ps $< .05$), whereas her scores for the two hard conditions did not differ significantly from chance (calculated as above, Fisher's Exact Test, one-tailed, ps $> .05$). In sum, neither of the two agrammatic Broca's aphasics showed evidence for a selective impairment for grammaticality judgment of sentences involving XP traces: one was at chance on all conditions, and the other performed well on easier sentences but at chance on more difficult sentences, regardless of whether or not the judgments depended on XP traces.

We also calculated the correlation between overall grammaticality judgment scores and two independent comprehension measures (Figure 5). Grammaticality judgment performance was quite highly correlated with performance on the screening test for agrammatic comprehension described above ($r = .66$, $p = .0008$) and was even more highly correlated with WAB comprehension subscores ($r = .71$, $p = .0002$). To confirm that these correlations do not just reflect some more general nonspecific impairment factor, we also calculated the correlation of grammaticality judgment with WAB fluency and found it to be not significant ($r = .29$, $p = .20$).

Aphasic Patients Grouped by Lesion Site

In the final set of analyses, patients were grouped according to the characteristics of their lesions. Lesion data were available for 21 of the 22 patients, comprising digital reconstructions for 16 patients and CT or MRI scans for 5. We first considered a possible role for the posterior inferior frontal gyrus (pIFG, Broca's area, Brodmann's areas 44 and 45), since it is widely held that

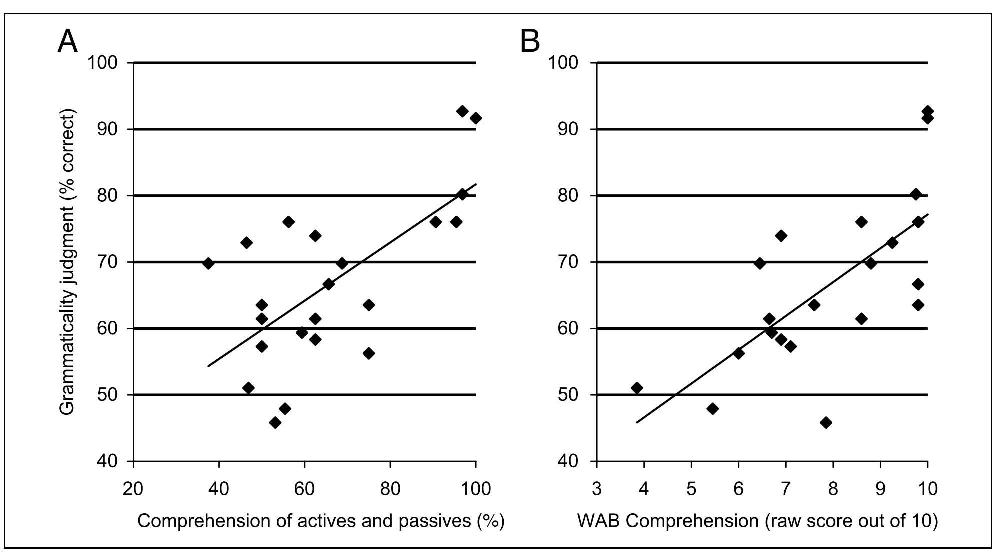

Figure 5. Scatter plots showing the performance of individual patients on grammaticality judgment as compared to (A) comprehension of actives and passives and (B) WAB comprehension subscores.

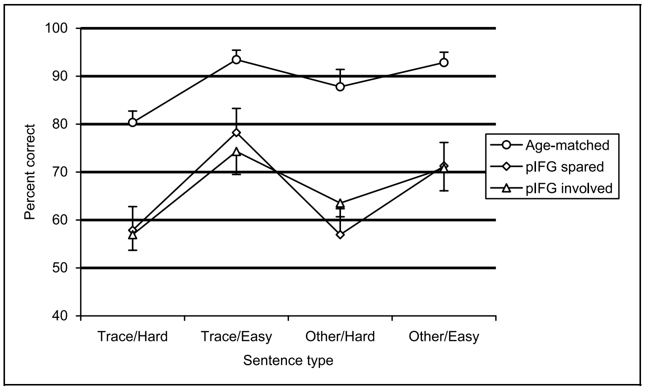

Figure 6. Grammaticality judgment performance of patients grouped according to whether their lesions included the pIFG and age-matched control subjects across the four sentence types.

this area is important for syntactic processing (e.g., Grodzinsky, 2000; Stromswold, Caplan, Alpert, & Rauch, 1996). Patients were divided into two groups: those whole lesions included all or part of the pIFG and those whose lesions completely spared this area. Figure 6 depicts the performance of these groups relative to age-matched controls.

A repeated-measures ANOVA revealed a main effect of group, $F(2,32) = 16.8$, $p < .0001$. The controls (89%) were significantly better than both the patient groups, but the patients with pIFG lesions (66%) did not differ significantly from those without pIFG lesions (66%). There was also a main effect of sentence type, $F(3,96) = 40.9$, $p < .0001$.

The interaction of group by sentence type was significant, $F(6,96) = 2.31$, $p = .048$. To determine the source of this interaction, we ran sub-ANOVAs comparing the two patient groups to controls, as well as a third sub-ANOVA comparing the two patient groups directly. When the patients with lesions involving the pIFG were compared to controls, there was a main effect of group, $F(1,24) = 30.5$, $p < .0001$, and a main effect of sentence type, $F(3,72) = 25.6$, $p < .0001$, but no interaction, $F(3,72) = 0.681$, $p = .55$. When the patients whose lesions spared the pIFG were compared to controls, there was a main effect of group, $F(1,21) = 24.3$, $p < .0001$, a main effect of sentence type, $F(3,63) = 31.1$, $p < .0001$, and a significant interaction of group by sentence, type, $F(3,63) = 5.14$, $p = .0051$. As can be seen in Figure 6, this interaction reflects the fact that the patients with lesions sparing the pIFG did relatively better on the trace/easy condition, and relatively worse on the other/hard condition. However, they were still severely impaired on all sentence types, suggesting that neither syntactic processing in general nor the processing of XP traces is strictly localized to the pIFG. When the two lesion-defined groups were compared directly, there was a main effect of sentence type, $F(3,57) = 26.5$, $p < .0001$, but there was neither a main effect of group, $F(1,19) = 0.0032$, $p = .96$, nor was there an interaction of group by sentence type, $F(3,57) = 1.66$, $p = .20$. This underscores the fact that the status of the pIFG had no effect on the overall severity of deficits in grammaticality judgment and little effect on the profile obtained across sentence types.

Given that damage to the pIFG did not appear to be specifically associated with any aspect of the task, we conducted a voxel-based lesion-symptom mapping (VLSM) analysis (Bates et al., 2003) to determine whether any other areas were more reliably associated with deficits in any aspect of task performance. VLSM involves carrying out statistical analyses of the relationship between tissue damage and behavior on a voxel-by-voxel basis and plotting the resultant statistics as color maps that depict the degree of behavioral involvement for each voxel. One of the primary advantages of VLSM is that it analyzes the relationship between continuous behavioral data and continuous lesion extents without the need for any cutoffs to be stipulated based on behavior or lesion site. As noted above, digital lesion reconstructions were available for 16 patients, and these analyses were based on these patients only. A t test was performed at each voxel, comparing the scores for patients whose lesions included that voxel to the scores for patients whose lesions spared that voxel.

Figure 7A, B and C shows maps for overall performance, performance on trace conditions, and performance on other conditions, respectively. The same area emerges as most reliably associated with poor performance on all three measures: a region encompassing parts of the posterior superior temporal gyrus (pSTG), the posterior superior temporal sulcus (pSTS), and the posterior middle temporal gyrus (pMTG). This is the approximate location of Wernicke's area. The similarities among the three maps reflect the fact that performance on different sentence types tends to be highly correlated. We made a second series of maps excluding 3 of the 16 patients who were left-handed, and similar results were obtained.

To confirm the importance of this posterior temporal region, we next divided patients into two groups based on whether their lesions included all or part of the pSTG, pSTS, or pMTG. Eighteen patients were included in this analysis; three were excluded as it was unclear from their scans whether or not this area was compromised. The results are shown in Figure 8. A repeated-measures ANOVA revealed a main effect of group, $F(2,29) = 22.9$, $p < .0001$. Unlike in the pIFG analysis reported above, here, all three groups differed significantly from one another (Tukey's HSD): controls (89%) scored higher than patients with lesions sparing the posterior temporal region (74%), who in turn scored higher than patients with posterior temporal lesions (60%), confirming that posterior temporal damage

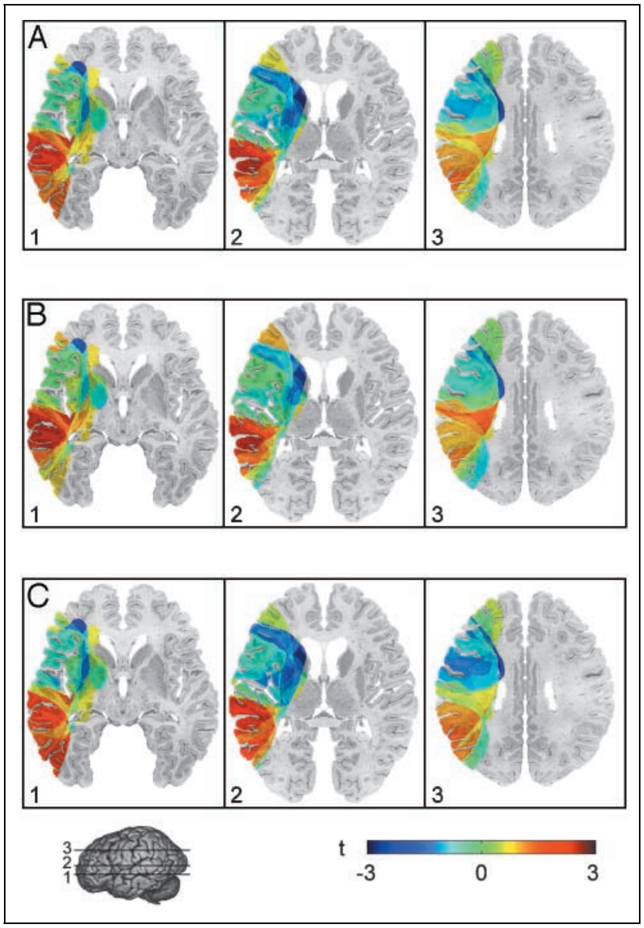

Figure 7. Axial VLSM displays showing the extent to which damage to each voxel was associated with task deficits. The values displayed at each voxel are *t* statistics (df = 14) comparing the patients lesioned at that voxel to the patients intact at that voxel. High *t* scores top the scale in red, indicating areas where damage led to significant deficits in task performance. Voxels denoted in blue reflect negative *t* scores, which arise when patients with lesions to those voxels performed better than those who had lesions elsewhere. Voxels that are not color-coded were not damaged in any of the patients in our sample. The behavioral measures displayed are (A) overall grammaticality judgment performance, (B) performance on structures where the contrast depends upon an XP trace (average of the trace/hard and trace/easy conditions), and (C) performance on structures where the contrast depends on other grammatical principles (average of the other/hard and other/easy conditions). The lateral view in the bottom left shows the approximate locations of the axial slices shown; this is not the same brain as shown in the slices, so the slice locations indicated are only approximate.

causes more severe deficits in grammaticality judgment than damage to other areas compromised in our patient sample. There was also a main effect of sentence type, $F(3,87) = 35.2$, $p < .0001$, but there was no interaction of group by sentence type, $F(6,87) = 1.56$, $p = .18$. This suggests that although patients with posterior temporal lesions are the most impaired in grammaticality judgment, they are not differentially impaired on any particular sentence type.

DISCUSSION

This study demonstrates that aphasic patients are usually substantially impaired in their ability to judge the grammaticality of sentences. However, there is little evidence that deficits are restricted to particular sentence types, regardless of how patients are classified. In particular, no patient groups were selectively impaired on sentences involving traces of maximal projections, contrary to the claims of Grodzinsky and Finkel (1998). When patients were grouped based on WAB classification, the only differential impairment found was with anomic patients, who performed somewhat more poorly on difficult sentences not involving XP traces, and somewhat better on easier sentences involving XP traces. Broca's aphasics and Wernicke's aphasics did not differ from age-matched controls in terms of their pattern of performance across the sentence types tested: Although they were severely impaired relative to controls, all sentence types were affected equally. When patients were identified based on an "agrammatic" comprehension profile in which comprehension of active sentences was above chance whereas comprehension of passives was at chance, there was likewise no evidence for a selective impairment of sentences involving XP traces nor for a selective sparing of sentences not involving XP traces. Finally, when patients were grouped based on lesion location, we found that patients with lesions involving the pIFG (Broca's area) did not perform any worse than patients whose lesions spared the pIFG, nor did their relative performance across sentence types differ from age-matched controls. Patients with lesions sparing the pIFG had a profile that resembles that of the anomic group, with worse-than-expected performance on difficult judgments not involving XP traces, and better-than-expected performance on easier judgments involving XP traces. A VLSM analysis revealed that a posterior temporal region comprising the pSTG, pSTS, and

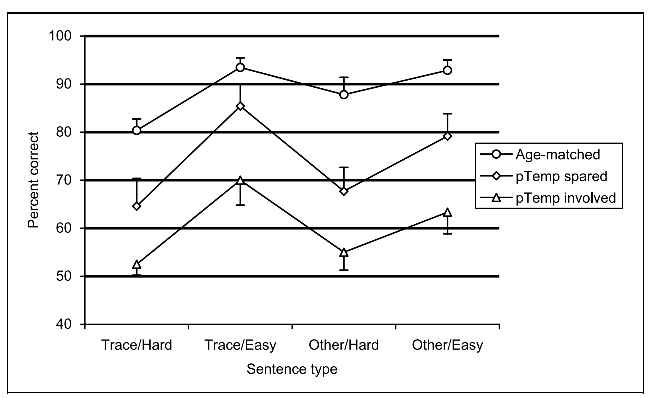

Figure 8. Grammaticality judgment performance of patients grouped according to whether their lesions included the posterior temporal region identified in Figure 7 (pTemp) and age-matched control subjects across the four sentence types.

pMTG was most reliably associated with deficits in grammaticality judgment. However, damage to this area did not affect any sentence type differentially.

There are three main respects in which the results obtained in this study are surprising: first, the fact that aphasic patients achieved such low scores in the grammaticality judgment task, in light of prior studies suggesting that grammaticality judgment is relatively intact in agrammatic aphasia (e.g., Linebarger et al., 1983); second, the fact that the relative performance pattern across sentence types is in most cases indistinguishable from that of normal controls, which contradicts claims that deficits are restricted to particular grammatical structures (e.g., Grodzinsky & Finkel, 1998; Mauner et al., 1993); and third, the finding that a posterior temporal region (approximately Wernicke's area) is more reliably implicated in poor performance than the pIFG (Broca's area), which is more often held to be particularly important for syntactic processing. We will discuss each of these findings in turn.

In this study, Broca's aphasics averaged 65% across all sentence types, Wernicke's aphasics averaged 58%, and anomic aphasics averaged 77%. This can be compared with the performance of age-matched controls who averaged 89%. Clearly, all aphasic groups have substantial deficits, even anomics, some of whom appear relatively unimpaired in naturalistic language use. Most other studies of grammaticality judgment have tested Broca's aphasics. The mean of 65% obtained by Broca's aphasics in our study can be compared to totals from studies such as Linebarger et al. (1983), where four Broca's aphasics averaged 82% correct (calculated from their Figure 1), Wulfeck and Bates (1991), where five Broca's aphasics obtained an average of 77% correct (calculated from their Table 1), and Grodzinsky and Finkel (1998) where four Broca's aphasics averaged 73% across conditions (calculated from their Table 2). Of course, each of these studies has used substantially different stimuli and somewhat different methods, so these disparate results are not unexpected. The subtlety of many of the judgments in the present study no doubt contributed to the relatively low overall scores observed. However, it does appear clear that with sufficiently difficult stimuli, substantial judgment deficits can be revealed. It is also noteworthy that we found quite high correlations of grammaticality judgment performance with comprehension measures, suggesting that comprehension and judgment tasks are probing receptive syntactic mechanisms in similar ways. The relative intactness of grammaticality judgment abilities has sometimes been taken to support a model of agrammatic aphasia in which linguistic competence is intact but performance is compromised (e.g., Linebarger et al., 1983). The substantial impairments observed in this study and the correlations with comprehension measures suggest that this view may need to be revised. One consideration is that in models of language where the competence/performance distinction is taken seriously, imperfect performance of normal subjects on judgment tasks leads inexorably to the conclusion that grammaticality judgment is itself a performance (Schütze, 1996), in which case grammaticality judgment tasks do not allow direct inferences about competence. Another possibility is that the distinction between competence and performance, while undoubtedly crucial for the study of natural language grammars, has little explanatory value in understanding aphasia. If knowledge of grammar is a kind of procedural knowledge (e.g., Ullman, 2001), then we might expect based on better-understood types of procedural knowledge, such as motor learning and perceptual learning that neural networks underlying the ability to carry out the task (''competence'') are the very same networks that subserve performance of the task (Karni et al., 1998). If this is true of the neural organization of language, then damage to these networks will inevitably affect both competence and performance, to the extent that they can be distinguished.

The second main finding is that aphasic groups defined in various ways tend to have performance profiles across sentence types that closely resemble the pattern obtained with normal controls. This contradicts the claim of Grodzinsky and Finkel (1998) that agrammatic Broca's aphasics, and possibly Wernicke's aphasics, have impairments that are specific to structures where the judgments crucially depend upon XP traces. By testing a wider range of sentence types, we found that there are some judgments involving XP traces that are performed relatively well by all aphasic groups [e.g., examples (10) and (11)] and there are some contrasts that do not involve XP traces yet which pose great difficulties for aphasic patients [e.g., examples (13) and (14)]. Patients' deficits were never observed to be restricted to particular classes of sentences. In most of the analyses carried out, there was no interaction of group by sentence type, indicating that any difficulties aphasic patients have with one sentence type over another are experienced equally by control subjects. As noted above, Grodzinsky and Finkel (1998) claim that age-matched controls obtained above 95% correct on all sentence types in their study, but we failed to replicate this result, obtaining results as low as 60% on the same sentence types used in their study. One possibility is that subjects in their study were inadvertently cued to the intended judgments by the experimenters. Although the procedure for the control testing is not reported, when the aphasic patients were tested, the stimuli were read by the experimenters. This is not an ideal procedure, especially given the difficulty of reading ungrammatical sentences with natural intonation. For this reason, the stimuli in our experiment were recorded in advance by an experienced phonologist who was blind to the conditions and to the hypotheses under consideration and presented in a computer-controlled procedure in which potential effects of cueing are minimized.

It is likely that some, or even many, patients may have comprehension deficits that do differentially affect particular syntactic structures (Caplan & Hildebrandt, 1988). However, what the present study and many others (e.g., Caramazza, Capitani, Rey, & Berndt, 2001) suggest is that although this may be true for single patients, when patients are grouped, group performance cannot be neatly described using linguistic or psycholinguistic constructs. In fact, in grammaticality judgment, there is little evidence for any differential impairments; patients were not even especially impaired on the more difficult sentences, relative to normal controls. This clearly contrasts with aphasics' performance on sentence comprehension tasks, where there is much evidence for differential impairments across sentence types (e.g., Dick et al., 2001; Grodzinsky, 2000; Schwartz et al., 1980; Caramazza & Zurif, 1976). Why does grammaticality judgment differ from comprehension in this respect? There are many models that can explain why some sentences are comprehended more easily than others. For instance, in the competition model (MacWhinney & Bates, 1989), it is argued that morphology is a "weak link" in language processing. When morphological cues are compromised, patients can still succeed on sentences with canonical word order, but they fail on sentences with less frequent word orders, since these sentences crucially depend on morphological cues for their interpretation (Dick et al., 2001). An approach along these lines can account for many of the findings regarding sentence comprehension in aphasia, which have been reported in the literature (Kay, 2000). The situation with grammaticality judgment is quite different, perhaps most importantly because there is no equivalent to canonical word order, that is, there is no default response available based on a more salient property of the sentence such as the order of its major constituents. Rather, almost all sentences contain overt morphological elements and other items of relatively low salience that must all be assessed correctly for grammaticality to be determined. This may be the main reason that dramatic interactions with patients performing relatively worse than normals on difficult sentences are not observed.

The question still remains as to why the "easy" sentences in the present study are easy and why the "hard" sentences are hard. It is important to note that the explanation should be one that applies to language processing in both normal and aphasic patients, given the failure to find group by sentence type interactions in most cases. Our study was not designed to answer this question, but we can suggest that almost all of the violations that subjects found most difficult to detect involved unstressed elements, generally with reduced vowels. For instance, rejection of *that-trace* violations depends upon an unstressed complementizer *that*, and rejection of the most difficult illicit auxiliary misplacements ("Could have she brushed her teeth?") requires sensitivity to the unstressed and reduced auxiliary *have*.

It is indisputable that deficits in aphasia can provide evidence about brain regions important for particular aspects of linguistic processing, for instance the role of the anterior insula in coordinating speech (Dronkers, 1996). However, syntactic processing has been much more difficult to tie to any particular neural location. The present study is consistent with several large-scale studies that have failed to find any evidence for selective deficits in syntactic processing specific to particular lesion sites (Dick et al., 2001; Caplan, Hildebrandt, & Makris, 1996; Caplan & Hildebrandt, 1988). Although our VLSM analysis revealed that damage to a posterior temporal region was most reliably associated with poor performance on the grammaticality judgment task, several patients with lesions that completely spared this area (generally left frontal lesions) were also severely impaired. Our results suggest that neural resources important for syntactic processing are probably distributed throughout the left peri-sylvian region (with possible support from the right hemisphere too). This appears to be the case for syntactic processing in general as well as more specific types of computation, such as the assessment of structures involving traces of maximal projections. There is, at present, no evidence that this linguistic concept is reflected at all in the neural organization of language. A recent fMRI study found that grammaticality judgment of sentences both with and without movement of phrasal constituents activated both Broca's and Wernicke's areas (Wartenburger, Heekeren, Burchert, De Bleser, & Villringer, 2003), a result that is consistent with our findings. Another area that appears based on lesion studies to play an important role in syntactic comprehension is the anterior superior temporal gyrus (Dronkers, Wilkins, van Valin, Redfern, & Jaeger, 1996).

Our results do not distinguish between a model in which different areas are specialized for different aspects of syntactic comprehension, but there is considerable individual variation in terms of the location of the particular areas (Caplan & Hildebrandt, 1988), or a model in which syntactic comprehension is distributed throughout the left peri-sylvian region, with little differentiation between particular areas (see Caplan et al., 1996, for discussion). However, the data presented are not compatible with a model in which computations involving XP traces are made exclusively in the pIFG (Grodzinsky & Finkel, 1998; Grodzinsky, 2000). As we have indicated above, some aspects of linguistic processing, such as motor planning for speech, are much more consistently localized than mechanisms for syntactic comprehension. Our results thus add to the growing consensus that the overall neural organization for language involves both localized and distributed features (Dick et al., 2001; Caplan et al., 1996; Damasio, 1992).

METHODS

Participants

Twenty-two patients diagnosed with aphasia participated in the study. Patients were recruited from Veterans' Administration Medical Centers and newspaper advertisements in San Diego, CA, or Martinez, CA, and were paid for their participation. All patients had left hemisphere vascular lesions and clinically diagnosed aphasia and were at least 1 year post onset at the time of testing. All were native English speakers with normal or correct-to-normal vision and hearing. Exclusionary criteria included diagnosed or suspected hearing difficulties, dementia, head trauma, tumors, or multiple infarcts. The age range of the patients was 51–83, with a mean of 67 years. Eighteen were right-handed and four were left-handed.

The patients were classified with the WAB (Kertesz, 1982), administered by a trained speech pathologist. By the WAB criteria, six patients were classified as Broca's aphasics, eight as Wernicke's aphasics, and eight as anomic aphasics. CT or MRI scans were available for 21 of the 22 patients. The remaining patient had a neurological report verifying that he had a single left-hemisphere infarction, but no brain scan. Sixteen of the 21 scans were digitally reconstructed onto 11 axial template slices from an atlas (DeArmond, Fusco, & Dewey, 1976) by a board-certified neurologist, using a computer program developed at the VA Northern California Health Care System (Frey, Woods, Knight, Scabini, & Clayworth 1987).

Twenty-six college-aged students (aged 18–36, mean 21) and 14 older people from the same communities and backgrounds as the aphasic patients (aged 54–79, mean 68) also participated in the study. Many of the older controls were spouses of patients in the study.

The study was approved by the VA Northern California Health Care System and UCSD Human Research Protection Programs. Informed consent was obtained from all subjects prior to their participation.

Materials

Aphasic patients were tested on 96 sentences. There were 24 sentences in each of four conditions: trace/hard, trace/easy, other/hard, and other/easy. Each of the conditions contained two sentence types. In each condition, there were 6 grammatical and 6 ungrammatical sentences of each of the two types, for a total of 24. Example sentences were shown above in Table 1. The other stimuli were closely based on these, generally, with variety being introduced by varying the noun phrases and verbs. Where possible, noun phrases and verbs were assigned randomly to grammatical and ungrammatical versions of sentences.

Both control groups were also presented with four additional sets of 12 sentences (6 grammatical and 6 ungrammatical), two in the trace/hard condition, and two in the other/easy condition. These sentences, like the others in the trace/hard and other/easy conditions, were closely modeled on those used by Grodzinsky and Finkel (1998). These additional sets of sentences were not tested on patients, as we wanted to keep the experiment short enough to be performed in a single sitting without overly taxing the patients.

The sentences were recorded in a soundproof booth at UCLA by an experienced phonologist, who was selected for his clear enunciation, unmarked accent, and expertise in phonology and phonetics. He was blind to the conditions involved in the experiment and was asked to read the grammatical sentences normally and the ungrammatical sentences with as normal an intonation contour as possible. Several independent judges listened to the sentences to verify the naturalness of the intonation, and a few were subsequently re-recorded. The sentences were digitized at 22 kHz and edited with SoundEdit. The experiment script was prepared with PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993).

Procedure

About half of the aphasic patients and most of the age-matched control subjects were tested at their homes by one of the authors (A.P.S.) and an experienced speech pathologist. Other patients, and some age-matched controls, were tested at UCSD or the Veterans' Administration Medical Center in Martinez. All of the college-aged control subjects were tested at UCSD.

Participants sat in front of a computer (Apple iBook), a pair of speakers (Yamaha YST-M7), and a standard PsyScope button box. The nature of the grammaticality judgment task was explained by the researchers. To explain the task, we used a script provided in the appendix to Grodzinsky and Finkel (1998), modified only very slightly to meet our own stylistic preferences. In brief, participants were instructed to reject sentences that either ''sound funny'' or ''don't make any sense,'' and they were encouraged not to see the experiment as a test of their abilities or any prescriptive grammar they might have been taught.

Eight warm-up trials were followed by 96 trials for patients or 144 for controls. Feedback was given during the warm-up trials, and it was ensured that participants understood the task. After the warm-up, trials were presented in random order. A trial consisted of the following. After a 2-sec wait during which a crosshair was displayed, the sentence to be judged appeared centered on the screen in large, clear type. Then, 0.5 sec after the sentence appeared, the recording of it was played. Subjects could respond at any time by pressing the green button on the button box if the sentence sounded okay or the red button if it sounded bad. There was also a green smiling face attached to the green button and a red frowning face attached to the red button. Subjects were permitted to ask for as many repeats as they desired, and most took the opportunity to do so several times. They could request a repeat either by pressing the third, yellow button on the button box, or by indicating to the experimenter. There was no timeout; subjects were instructed to guess if necessary.

When testing patients, the researchers sat behind the subjects so that they could not inadvertently provide any cues as to the grammaticality of sentences. Encouraging feedback was provided periodically to patients, but it was never related to success or failure on any particular trial. Control subjects were generally given minimal feedback after the warm-up trials and only monitored for general attention and alertness.

Statistical analysis was performed with JMP. Repeated measures ANOVAs were Geisser–Greenhouse corrected where appropriate. Items analyses were performed but are not reported as they were generally consistent with subjects analyses. Lesion analysis was carried out with VLSM software (Bates et al., 2003). This software operates on lesion files in the popular ANALYZE image format and is freely available on-line at http://crl.ucsd.edu/vlsm.

Acknowledgments

This research was supported by a grant to Elizabeth Bates, ''Cross-Linguistic Studies in Aphasia'' (NIH/NIDCD 2 R01 DC00216). We thank Elizabeth Bates for providing funding, access to patients, and testing facilities; Nina Dronkers for providing patients and testing facilities; Bob Knight for lesion reconstructions; Eva Schleicher for assistance with running control subjects; Suzanne Moineau and Carl Ludy for assistance with testing aphasic patients; Bruce Hayes for reading stimulus sentences; Bob Buffington for assistance with experiment scripts; and Fred Dick and Tim Stowell, along with all of the people mentioned already, for helpful discussions and comments. We also thank the patients who gave their time to participate in the study.

Reprint requests should be sent to Stephen M. Wilson, Neuroscience Interdepartmental Program, University of California, Los Angeles, 1320 Gonda Center, 695 Young Drive South, Los Angeles, CA 90095-1761, USA, or via e-mail: stephenw@ucla.edu.

REFERENCES

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience, 6,* 448–450.

Beretta, A., Piñango, M., Patterson, J., & Harford, C. (1999). Recruiting comparative crosslinguistic evidence to address competing accounts of agrammatic aphasia. *Brain and Language, 67,* 149–168.

Blackwell, A., & Bates, E. (1995). Inducing agrammatic profiles in normals: Evidence for the selective vulnerability of morphology under cognitive resource limitation. *Journal of Cognitive Neuroscience, 7,* 228–257.

Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension.* Cambridge: MIT Press.

Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain, 119,* 933–949.

Caramazza, A., Capitani, E., Rey, A., & Berndt, R. S. (2001). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. *Brain and Language, 76,* 158–184.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language, 3,* 572–582.

Chomsky, N. (1995). *The minimalist program.* Cambridge: MIT Press.

Chomsky, N. (2000). Linguistics and brain science. In A. Marantz (Ed.), *Image, language, brain: Papers from the first mind articulation symposium* (pp. 13–28). Cambridge: MIT Press.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers, 25,* 257–271.

Damasio, A. R. (1992). Aphasia. *New England Journal of Medicine, 326,* 531–539.

DeArmond, S. J., Fusco, M. M., & Dewey, M. M. (1976). *Structure of the human brain: A photographic atlas.* New York: Oxford University Press.

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.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature, 384,* 159–161.

Dronkers, N. F., Wilkins, D. P., van Valin, R. D. Jr., Redfern, B. B., & Jaeger, J. J. (1996). A reconsideration of the brain areas involved in the disruption of morphosyntactic comprehension. *Brain and Language, 47,* 461–463.

Frey, R. T., Woods, D. L., Knight, R. T., Scabini, D., & Clayworth, C. C. (1987). Defining functional areas with averaged CT scans [Abstract]. *Society for Neuroscience Abstracts, 13,* 1266.

Goodglass, H. (1993). *Understanding aphasia.* San Diego, CA: Academic Press.

Grodzinsky, Y. (1986). Language deficits and the theory of syntax. *Brain and Language, 27,* 135–159.

Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences, 23,* 1–21.

Grodzinsky, Y., & Finkel, L. (1998). The neurology of empty categories: Aphasics' failure to detect ungrammaticality. *Journal of Cognitive Neuroscience, 10,* 281–292.

Grodzinsky, Y., Piñango, M. M., Zurif, E., & Drai, D. (1999). The critical role of group studies in neuropsychology: Comprehension regularities in Broca's aphasia. *Brain and Language, 67,* 134–147.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1998). The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 861–868.

Kay, P. (2000). Comprehension deficits of Broca's aphasics provide no evidence for traces. *Behavioral and Brain Sciences, 23,* 37–38.

Kertesz, A. (1982). *Western aphasia battery.* New York: Grune & Stratton.

Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13,* 361–392.

Lu, C.-C., Bates, E., Li, P., Tzeng, O., Hung, D., Tsai, C.-H., Lee, S.-E., & Chung, Y.-M. (2000). Judgements of grammaticality in aphasia: The special case of Chinese. *Aphasiology, 14,* 1021–1054.

Lukatela, K., Crain, S., & Shankweiler, D. (1988). Sensitivity to inflectional morphology in agrammatism: Investigation of a highly inflected language. *Brain and Language, 33,* 1–15.

MacWhinney, B. & Bates, E. (Eds.). (1989). *The crosslinguistic study of sentence processing.* Cambridge: Cambridge University Press.

Mauner, G., Fromkin, V. A., & Cornell, T. L. (1993). Comprehension and acceptability judgments in agrammatism: Disruptions in the syntax of referential dependency. *Brain and Language, 45,* 340–370.

Rizzi, L. (1990). *Relativized minimality.* Cambridge: MIT Press.

Schütze, C. T. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology.* Chicago: University of Chicago Press.

Schwartz, M. F., Linebarger, M. C., Saffran, E. M., & Pate, D. S. (1987). Syntactic transparency and sentence interpretation in aphasia. *Language and Cognitive Processes, 2,* 85–113.

Schwartz, M. F., Saffran, E. M., & Marin, O. S. (1980). The word order problem in agrammatism: I. Comprehension. *Brain and Language, 10,* 249–262.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language, 52,* 452–473.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience, 2,* 717–726.

Wartenburger, I., Heekeren, H. R., Burchert, F., De Bleser, R., & Villringer, A. (2003). Grammaticality judgments on sentences with and without movement of phrasal constituents—An event-related fMRI study. *Journal of Neurolinguistics, 16,* 301–314.

Wulfeck, B., & Bates, E. (1991). Differential sensitivity to errors of agreement and word order in Broca's aphasia. *Journal of Cognitive Neuroscience, 3,* 258–272.

Wulfeck, B., Bates, E., & Capasso, R. (1991). A crosslinguistic study of grammaticality judgments in Broca's aphasia. *Brain and Language, 41,* 311–336.