Available online at www.sciencedirect.com
SCIENCE DIRECT®

Brain and Language 96 (2006) 157–170

Brain and Language

www.elsevier.com/locate/b&l

Reply

# The variability debate: More statistics, more linguistics ☆

Dan Drai [a], Yosef Grodzinsky [b,c,*]

[a] Weizmann Institute, Israel
[b] McGill University, Canada
[c] Tel Aviv University, Israel

Accepted 19 May 2005
Available online 2 November 2005

**Abstract**

We respond to critical comments and consider alternative statistical and syntactic analyses of our target paper which analyzed comprehension scores of Broca's aphasic patients from multiple sentence types in many languages, and showed that *Movement* but not *Complexity* or *Mood* are factors in the receptive deficit of these patients. Specifically, we do the following: (a) We show how group analyses such as ours are valid and critically important, and then discuss apparent statistical discrepancies between our analysis and others'. (b) We provide new syntactic arguments in support of our decision to categorize passive sentences in German and Dutch as −*Movement* in the context of Broca's aphasia. These arguments serve to underscore an intriguing correlation between German/Dutch and English: On the one hand, Scope Freezing is found in the former where the latter allows scope ambiguity. On the other hand, Broca's aphasic patients successfully comprehend German/Dutch passive, but fail in English. (c) We reanalyze new data from Dutch and Italian passive, which point to new and potentially interesting cross-linguistic differences. Our current analyses are based on an addition of raw scores from 62 new patients to the existing data base of 69 Broca's aphasic patients. We conclude that while aphasic performance is indeed variable, the group results have distinct statistical and linguistic structure.


*Keywords:* Broca's aphasia; Variability; Quantitative analysis; $\beta$-Distribution; Syntax; Movement; Cross-linguistic variation

## 1. Preliminaries

In Drai and Grodzinsky (2005), our target article, we presented the results of quantitative explorations into a large data set of comprehension scores from 69 Broca's aphasics. These results lend empirical support to the following claims:

1. That the methods we developed (i.e., the construction of the database, the presentation of performances by confidence intervals built on a binome, and the use of the $\beta$-distribution as a basis for hypothesis testing) open new vistas on the data that no single-case approach makes possible.
2. That the perspective we offer enables the quantification of variation.
3. That our analytic methods discover robust structure in the data set, inherent in relative success in the comprehension of sentences that do not involve syntactic movement, and failure when sentences are derived by movement.
4. That other partition methods fail to detect structure in the data set: neither "active/passive" nor "high/low sentence complexity" are viable distinctions for deficit analyses in Broca's aphasia.

DOI of original article: 10.1016/j.bandl.2004.10.016.
☆ Supported in part by Canada Research Chairs, Canada Foundation for Innovation, and a McGill VP-Research internal grant to Y.G. D.D.'s research at the Weizmann Institute of Science is supported by the Golda and Dr. Yehiel Shwartzman and Sara and Haim Medvedi Families Postdoctoral Fellowship. We thank Roelien Bastiaanse, Kathlyn Clark, Na'ama Friedmann, Ineke van der Meulen, Andrea Santi, and Christiane Ulbrich for their help.
* Corresponding author. Fax: +1 514 398 7088.
*E-mail address:* Yosef.Grodzinsky@mcgill.ca (Y. Grodzinsky).

0093-934X/$ - see front matter 
doi:10.1016/j.bandl.2005.05.004

We concluded that, despite obvious variability of the type that exists in virtually every realm where real-life numerical data are discussed (at least in biology), the explorations we conducted in the large data set we created lend support to a view of Broca's aphasia as a deficit in syntactic movement, along the general lines of the trace-deletion hypothesis (TDH).

Our critics have responded with several counterarguments. Some have challenged our quantitative assumptions and analyses, others our linguistic description as well as the empirical evidence we used. Two alternative accounts were proposed and one additional database was created and analyzed. In what follows, we make an attempt to use these criticisms in order to enhance our understanding of the current state of affairs and to create a somewhat clearer picture of the role Broca's region plays in linguistic analysis.

Here is how we proceed: In Section 2, we note the aspects of our analysis which are consensual (or at least unchallenged). In Section 3, we turn to the critique of our numerical analyses (raised by Caplan et al., by Toraldo & Luzzatti, and by Amunts and Willmes), and to alternative methods (and putative results) that Caplan et al. present. In Section 4, we discuss issues that pertain to the syntax of passive in Germanic languages, responding to linguistic objections raised by Bastiaanse and Zonnenfeld and de Bleser, Burchert, and Schwarz. In the same section, we also discuss the quantitative analysis of a study of active/passive in Dutch (Bastiaanse and Edwards, 2004). Section 5 focuses on group results from another two studies of Italian actives and passives, by Caramazza, Capasso, and Capitani (2005) and Luzzatti et al. (2000). Section 6 considers anatomical issues (Amunts and Willmes), and Section 7 discusses alternative linguistic analyses (proposed by Bastiaanse and Zonnenfeld and by Friedmann). The discussion opens way to a host of new empirical issues and we thus offer the readers access to our dataset, available as supplementary material on http://freud.tau.ac.il/~yosef1/. We do so in order to encourage further quantitative analyses of this dataset. Finally, we try to come up with an interim conclusion of the debate.

## 2. Where everyone agrees

### 2.1. Contextualizing the debate

The critique in several papers mixes issues that pertain to the analysis of the data set, with putative problems for the TDH. While these criticisms may all be true, it is important to separate the statistical from the linguistic discussion. To see why, let us put matters in perspective.

Theoretical proposals are always contextually dependent. When syntactic discussion of comprehension in aphasia began, the range of available data on sentence comprehension in aphasia was limited. The record contained results from few syntactic constructions, tested on a handful of aphasic patients. The first version of the TDH was formulated in this context and was thus based on results for a few patients whose comprehension skills were tested on actives, passives, and object relative clauses (Grodzinsky, 1984, 1986).

Neurolinguistics has since grown, and the data base expanded, naturally bringing about a more complicated picture. Linguistically, complex new data led to novel theoretical accounts, as researchers attempted to push the envelope of our understanding further (cf. Avrutin, 2001, for a recent review of several approaches). Clinically, scores for larger numbers of patients gradually accumulated (through a variety of experimental methods) and further complicated the picture. The difficulty to see clear patterns in these led some to believe that Broca's aphasia is a spurious generalization, and to argue that patients' performance levels vary boundlessly. It was mostly this latter issue that our study addressed. As no serious quantitative analysis existed, we decided to try.

Our decision to examine actives, passives, and relative clauses was motivated by pragmatic as opposed to conceptual considerations: we wanted to scope performance variation, and as these constructions have been repeatedly tested in Broca's aphasia over the years, they provided us with a good testing ground. Thus, we note that while the results are highly relevant to any account of Broca's aphasia (the TDH being no exception), the data that are available today go way beyond this limited and somewhat accidental set of constructions. Moreover, if we could rewrite history, we would probably prefer to study other sentence types in aphasia.

The discussion below touches on statistical, linguistic, clinical, and anatomical issues. We first discuss variation in the data set and then move on to attempts to account for these data.

### 2.2. Consensual issues

We begin by specifying uncontroversial claims: Everyone agrees that more data are better than less, and that group analyses allow a broader picture. Everyone also agrees (and in fact one reanalysis of the data reconfirms) that movement is the correct distinction in the realm of relative clauses (subject vs. object gap), while "complexity" (branching type) is not. Also unchallenged is the claim that syntactic movement sets types of active sentences apart from one another (i.e., base actives vs. scrambled or topicalized ones), and that comprehension scores of German/Dutch Broca's aphasic patients on passive sentences are significantly higher than those of their English counterparts. Empirically, this touches on a large portion of the dataset we constructed: It contains performance scores of 32 patients who participated in relative clause tests, 18 patients (speakers of non-English languages) who participated in movement tests in the domain of active sentences, and 27 English speaking patients and 24 Dutch and German patients who partook in the passive test. All these results, then, are beyond debate.

Disagreement, then, remains only on the relation between active and passive and its cross-linguistic manifestations. And, while the response to the critique will lead us to digress into statistical, syntactic, and experimental matters, it should be clear that the scope of the debate has narrowed down significantly: No other fact under discussion (or for that matter, no other relevant empirical fact that is currently available) has been challenged.

## 3. Statistics

This section has a somewhat technical character. For readers less interested in these intricacies, we provide an informal summary at its end. The quantitative approach we took fits a $\beta$-distribution to the group data for each condition, and tests for the significance of the differences in the resulting distributions among the relevant contrasts, thus demonstrating structure in the dataset. This approach, as Amunts and Willmes aptly realize (and commend), introduces a "second-level (random effects) model approach to the analysis of binary-choice success rate data." However, both Caplan et al. and Amunts and Willmes raise a number of issues; Caplan et al. further propose an alternative analysis, which in one case (the ***Mood*** contrast) leads to an empirical claim that is at odds with one of ours.

### 3.1. Amunts and Willmes

These authors begin with the correct observation that our method assumes "an identical success probability for all items within an item set per individual patient." Indeed, we make two assumptions about the studies we analyze: (i) each subject has a probability of performing correctly on a given binary-choice task, and may have a different probability of success for another task. (ii) The actual performances of the subjects on a task evidence the value of this probability parameter. Our approach evaluates the distribution (in the population of Broca's aphasics) of these probabilities on a given task.

Amunts and Willmes question the validity of the first assumption. They then bring up a technical point to the effect that the existence of a success probability for a patient on a task "seems to be a less tenable assumption for cumulative single-patient studies all using their own specific definition of the item sets employed. Furthermore, the assumption of a constant (positive) intra-item set correlation may be tenable for a litter (see Williams, 1988) but not so easily for a set of items answered by the same patient."

We disagree with both points: (a) all the studies use the same definition of the item sets employed, since these are defined in terms of their grammatical structure, so as to constitute minimal pairs for the contrast under investigation, (b) the assumption of intra-item set correlation for a litter intuitively says that the probability of success of a given process occurring repeatedly in an organism is determined by traits of this organism. Our assumption simply amounts to saying that a certain computation that occurs repeatedly in a given brain has a probability of success determined by the properties of that brain. There is nothing non-standard in these assumptions.

Amunts and Willmes finally argue that our implementation of the Williams (1975) approach should be replaced by a more robust variant. It seems that this suggestion is based on a misreading of Williams (1988).[1]

### 3.2. Caplan et al.

These authors also address analytic issues. Their first objection is related to the properties of the sampling through which we try to evaluate the distribution of the probability of success on a given task for the population of Broca's aphasic patients. They observe that sometimes we include scores of the same patient on two sentence types, thereby sampling correlated scores. They proceed to claim that it is not legitimate to ask whether the resulting distributions are significantly different. This claim is presented without

[1] Amunts and Willmes write: "More importantly, identity of the intra-item set correlation for the two item sets to be compared (e.g., +/− *Movement*) must be violated if the success rates in both sets are different, as Williams (1988, p. 306) has pointed out. He advocates the more robust maximum quasi-likelihood estimators." First, let us point out that the quoted paper (Williams, 1988) deals specifically with the fitting of a dose–response regression model in teratological studies, and it is in this with respect to this problem (dose–response fitting) that the concern is raised. Indeed Williams continues in the next section: "The primary aim of the statistical analysis may be to establish whether there is any evidence of difference between dose groups, rather than to fit a dose–response regression." (this is the case in our analysis), and he finally concludes, in this context that is the one relevant to our application, "If we assume [*equality of intra-litter correlation*] (...) the differences [*between the means of the β-distributions*] will be underestimated." What this means concretely is that if we assume equality of correlation structure between all doses we run the risk of underestimating the differences between the resulting distributions. But this is exactly in agreement with what we do in our numerical procedure, we do not assume a priori that for each subject the correlation between answers is invariant from one type of sentence to another. So, in brief, we do not see how the points discussed in Williams (1988) affect our analysis.

an argument, and is probably unwarranted, since it would imply that one can never study the distributions of repeated measures of a population under different sets of conditions. For instance, Caplan et al.'s view excludes the use of repeated measures ANOVA for the analysis of results of standard psychological experiments. Yet after making this claim, Caplan et al. proceed to raise the opposite concern and demand that repeated sampling under two conditions should exclusively be carried out on the *same* population: "Drai and Grodzinsky's Fig. 4D depicts different $\beta$-functions [...], but many more patients contribute to the data points for the [−***Movement***] sentences than to those for the [+***Movement***] sentences [...] If these patients are ones whose performance is overall very good, their performances on [+***Movement***] sentences might also be good, resulting in a quite different $\beta$-curve." The consequence of this claim would be that the patients tested on active sentences only scored higher than those tested for both active and passive. While this concern obviously contradicts the preceding one, it can fortunately be tested empirically: We thus computed the $\beta$-distribution for scores on active sentences for the patients who had been tested on actives only ("non-paired"), and compared it to the $\beta$-distribution for patients tested on both active and passive ("paired"). No difference is observed: "non-paired"—$\mu = 0.86$, $\sigma = 0.14$; "paired"—$\mu = 0.88$, $\sigma = 0.07$. We can safely conclude that this objection is misplaced.

This takes care of the negative part of Caplan et al.'s critique. But there is a more positive aspect to it: They propose and implement an alternative method of analysis, in which only subjects tested on both sides of a contrast are taken into account. They test each subject individually for performance difference on both sides of each contrast, and then test the resulting pool of $\chi^2$ scores against chance. While this strategy leads them to discard a sizeable portion of the data set, it leads them to empirical conclusions similar to ours in two of the three analyses they carry out (a significant ***Movement*** effect and no significant ***Complexity*** effect in relative clauses). Their last analysis isolates ***Mood*** from ***Movement*** in mono-clausal sentences, in order to test for a ***Mood*** effect, namely, to compare actives and passives in instances when ***Movement*** is neutralized. The comparison between the pooled $\chi^2$ scores and a random distribution reached significance, indicative of a ***Mood*** effect.

The idea to test this contrast is excellent. We therefore tested it through our method. The resulting test approaches significance: for $n = 19$: $\mu_{\text{active}} = 0.89$, $\sigma_{\text{active}} = 0.125$; $\mu_{\text{passive}} = 0.78$, $\sigma_{\text{passive}} = 0.15$; the probability for a difference between the two on the $\beta$-distribution-based hypothesis testing $p = .056$. The best fitting $\beta$-curves are given in Fig. 1 (active—perforated red line, passive—solid black line).

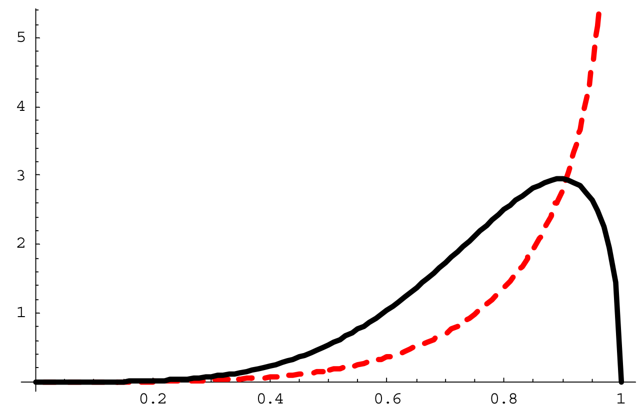


Fig. 1. $\beta$-Distributions for the Mood contrast according to Caplan et al.

Note that our $p$ value is different from Caplan et al.'s (who obtained $p = .0001$). This is related to the fact that our question is different from theirs: We model the overall distribution of probability of success on each branch of the contrast and then test whether these distributions are significantly different. Caplan et al., by contrast, try to see whether for each subject, the claim that ***Mood*** affects performance is a reasonable one, and then show that the pooled individual results obey a familiar pattern: performance on actives is generally better than passives.

The difference between the approaches can be illustrated thus: In a case in which most subjects do slightly better on one branch of a contrast than on the other, yet the difference between the success probability distributions of the two branches is not significant, Caplan et al.'s method would obtain an effect, whereas ours would not. In this sense, their approach might sometimes usefully exploit information that our approach would discard. However, there are opposite situations: our method enables precise framing of the range and extent of inter-subject variability, as well as useful group statistics. We compute sample means and variances, which enables us to create a picture of our population. For example, the (mostly −***Movement***) passive scores analyzed by Caplan et al. (solid black line in Fig. 1) after they discard data from "unpaired" patients, leads to a significantly higher mean than that of the +***Movement*** passive scores (Table 2 in the target paper). Namely, our result that German/Dutch passive is higher than the English one cannot be detected by Caplan et al.'s method.

In brief, when trying to exploit an empirical database such as ours, different analytic methods exploit and discard different types of information. The choice of method is guided by pragmatic considerations; ours followed from a desire to shed light on the variability debate, and show that in spite of the variability, group studies are a viable, in fact in some cases the only viable, approach.

Finally, Caplan et al. discovered a new structural relation among parts of the dataset we constructed—they proposed a way to partition the data which, as we concluded, results in a ***Mood*** effect. While this finding is potentially important, and still awaits an account, it hardly warrants Caplan et al.'s conclusion that the TDH is "refuted." The core predictions of the TDH are borne out: a robust movement effect in both mono-clausal and bi-clausal sentences, independent of ***Mood*** and ***Complexity.*** Moreover, the overall ***Mood*** effect obfuscates a subtle cross-linguistic difference: Unlike the other languages in the database, Dutch and German do not manifest a ***Mood*** effect. As the next section shows, the absence of a ***Mood*** effect follows from properties of the passive construction in these languages and suggests that the problem in front of us is more complicated than previously believed.

### *3.3. Toraldo and Luzzatti*

These authors accept our results, but make several critical points. First, they argue that despite the structure in the data that we demonstrated, there is still quite a lot of between-subject variability. To clarify our thinking about this remaining variability, they propose a distinction between a qualitative and a quantitative type. They argue that some variable results imply inter-syndromic differences (qualitative), whereas others are mere differences of quantity within the same syndrome (quantitative). Our data and theirs, they further argue, contains variability that we missed, which could be qualitative in nature. This may have happened because we used the $\beta$-distribution, which led us to spurious generalizations.

We certainly agree that there is still unexplained variation, and accept the distinction between the qualitative and quantitative types. It is clear that our data set contains only the categories we focused on. In our target paper, we pointed out that the dataset contains variability above and beyond the distinctions we uncovered. We also proposed a way to think about it - that in cases like the passive construction (which Toraldo and Luzzatti analyze) each patient has an individual probability parameter. Toraldo and Luzzatti propose to investigate the distribution of this parameter whithin one of the categories we considerer (passive), and we cannot agree more. It is for this reason that we made our data set public, so that other researchers can use it to make new discoveries.

Whether this additional variability is inter- and intra-syndromic is more difficult to tell. Medicine categorizes pathological phenomena, but within each category there are individual differences. But there is no a priori way to know whether our results represent a variable, yet single, syndrome, or more than one syndrome.

Toraldo and Luzatti express concern over the use of the $\beta$-distributions model, mentioning the fact that these distributions are unimodal. While this is true, it does not preclude using the distributions to discover multimodality, in fact this is exactly what we do when differentiating between the $+$***Movement*** and $-$***Movement*** performances, with the added benefit of an hypothesis testing framework for assessing the exact extent to which bi-modality is a more reasonable assumption than unimodality. They also observe that there may be a group of patients that perform at ceiling on passive. This may be true, yet in order to see whether or not it implies a new syndrome, much more work is needed: first, it is necessary to establish a quantitative method to test whether this performance pattern is indeed distinct from the rest of the group; second, if such a distinction is indeed found, we should then try and correlate this with other performances. The first step - establishing a distinction on quantitative ground - requires the reliance on an hypothesis testing framework, which will involve a family of distributions ($\beta$ or others). For the second step, one should think about other performances that are expected from patients who are at ceiling in passive, show (by $\beta$-tests and the like) that these performances also distinguish them from Broca's aphasic patients, and then try to establish numerical correlations that would build a profile of a new syndrome.

No one has thus far attempted such a daring move. The data base we now made public this makes the beginnings of such a project be possible. We certainly hope that someone would take the challenge.

### *3.4. Informal summary*

Let us summarize this section for the statistically less-proficient readers. Statistical objections to our paper regarded (1) the assumptions we made regarding the structure of our data set, (2) our analyses, which were said to discard data, and (3) our conclusions, as a new result—a ***Mood*** effect—was obtained. We argued that the objections in (1) are unjustified (or at least that some of them could be leveled against any empirical investigation in neuroscience); that (2) while every analysis discards some aspects of the data, ours provides the most comprehensive numerical picture among those under consideration, and that in any event, most effects in the dataset are sufficiently robust so that they hold across different analytic methods. Finally, we acknowledged (and replicated) the ***Mood*** effect, but pointed out that while being a new discovery, it does not diminish the force of our TDH-based claims, and in fact focusing on this last effect diverts attention from another new finding, that contrasts Dutch/German passive with the other languages, and to which we turn directly.

## 4. The structure of passive in Dutch and German

### 4.1. Scope ambiguity differences between English and German

The English sentence (fragment) in (1) has two meanings, while its German counterpart (2) only has one:

(1) Since he ***forgot*** to close *all the windows*...
(2) Weil er *alle Fenster* ***vergessen*** hat zu schließen...[2]
'Since he all windows(-ACC) forgotten has to close...'

English speakers judge (1) as true when the subject's negligence has the result that not all the windows were closed. That is, he may have remembered to close some of the windows, but not all of them ended up being closed, because he forgot to close some (i.e., with the predicate ***forget*** taking scope over the quantifier *all*—(***forget*** $\gg$ *all*)). This sentence is also judged true if the result is that *all* windows remained open; that is, if for all the windows it is true that he forgot to close them (i.e., with the quantifier *all* taking scope over the predicate ***forget***—(all $\gg$ ***forget***)). This exemplifies the well-known scope ambiguity phenomenon, which in (1) is putatively related to the existence of a negation operator within the verb ***forget***, a fact that turns this otherwise innocuous verb into a Scope Bearing Element (SBE).

Almost equally well known is the absence of this phenomenon in certain analogous German (and Dutch and Japanese) sentences: Unless special intonational devices are used (to obtain a focus reading), the sentence in (2) only entails that all the windows remained open and not a single one was closed—it only has the interpretation in which the relationship between *all windows* and ***forget*** corresponds to their surface order (namely, where the quantifier *alle* takes scope over ***vergessen***). This property of German is known as Scope Freezing (e.g., Frey, 1993; Krifka, 1998; Sauerland, 2003; Wurmbrand, 2003).

Scope ambiguities like (1) are typically accounted for by assuming an abstract operation that moves the English quantified expression *all the windows* over the verb ***forget.*** This operation is known as Quantifier Raising, or QR (cf. Fox, 2003 for a recent exposition). The reason why the German sentence lacks the "inverse scope" interpretation (that is, the one in which the linear ordering of the quantifier *alle* and the verb in the post-QR representation is the reverse of their surface order) is subject to debate (cf. Sauerland, 2003 for recent discussion).

Not that scope ambiguities are unavailable in German: they do exist, but overt movement is required to obtain them. Sentence (3), which differs from (2) only in that one SBE (the predicate ***forget*** with its embedded negation) moves overtly and crosses another (the quantified expression *all the windows*), is ambiguous, and just like the English (1) has two entailments—first, that some, but not all, the windows remained open, and second, that his forgetfulness pertained to each and every of them:

(3) Er ***vergaß*** *alle Fenster* ***t*** zu schließen
He forgot all windows (-ACC) to close
'He forgot to close all the windows'

The reason for the cross-linguistic difference may be somewhat of a mystery, but it is nonetheless clear that the availability of scope ambiguity in (3) and its absence in (2) correlate with overt syntactic movement. That is, the English-like ambiguity in (3) seems to become available as the verb moves leftward, crossing the quantified expression. Ambiguity between two SBEs in German is thus impossible in configurations such as (4a) which schematizes (2), but possible in (4b) that corresponds to (3):[3]

| | | |
|---|---|---|
| (4) a. | ...$SBE_1$... $\boldsymbol{SBE_2}$... | 1 meaning: (*all windows* $\gg$ ***forget***) |
| b. | ...$\boldsymbol{SBE_2}$...$SBE_1$...$\boldsymbol{t}_{SBE_2}$... | 2 meanings: (*all windows* $\gg$ ***forget***) (***forget*** $\gg$ *all windows*) |

Scope ambiguities in German thus correlate with certain movement operations. We can use their presence or absence (as evidenced by the availability or absence of certain meanings) as a diagnostic test for these movement types. Armed with this test, we now move on.

### 4.2. Passive in aphasia, and the English-Germanic contrast

Bastiaanse and Zonnenfeld and de Bleser, Burchert, and Schwarz argue that we misanalyzed passive in German and Dutch. They point out that we placed it in the −***Movement*** category, despite the widely accepted view that it is derived by movement. As a first step toward a reply, we

[2] The presence or absence of ambiguity depends on the case of *Alle Fenster*. When Nominative (i), the sentence is ambiguous; when Accusative (ii), it is unambiguous (see Wurdbrand, 2003 for discussion of such contexts, where further complexities arise, that need not concern us):

(i) *weil alle Fenster zu schließen vergessen* ***wurden***
since all windows (-NOM) to close forgotten **were**
'since they forgot to close all windows'
(ii) *weil alle Fenster zu schließen vergessen* ***wurde***
'since all windows(-ACC) to close forgotten **was**'.

[3] The presentation here is simplified, and ignores other complications, most notably the fact that the NP *alle Fenster* actually starts out as the object of *zu schließen.* It is not relevant to the present analysis because the verb *vergessen* starts to the right of the NP and moves leftward. As a consequence, both SBEs move leftward, maintaining the scopally relevant order between them, and hence the depiction in (4) can ignore these initial movements. Cf. Wurmbrand (2003).

note that our sin might actually have been greater: We analyzed not only passives as −***Movement***, but also *actives,* even though it is well known that their derivation, too, involves movement. Specifically, it is currently more-or-less uncontroversial that subjects, even in active sentences, move from the specifier position of VP to a higher surface position (cf. Koopman & Sportiche, 1991):

(5) **He**$_1$ [**t**$_1$ closed *the window*]

We ignored this fact, because this leftward movement from the left of the verb does not seem to interact with the comprehension deficit in Broca's aphasia (cf. Grodzinsky, 1995a, in press, for discussion of this point). Indeed, our critics rightly did not take us to task on it, seemingly realizing that we were only concerned with syntactic movement that is relevant to Broca's aphasia. What they do question is our decision to categorize Dutch passive (represented in the target paper as in (6)) as −***Movement***:

(6) (Door het meisje) wordt **de jongen** ◀ gekust
'(by the girl) was the boy ◀ kissed'

In our target paper, we used a leftward pointing arrow "◀" to annotate some sort of displacement, distinguishing it from another sort, an arrow pointing rightward "▶" which we represented as +***Movement***. We categorized the former, evident in Germanic passive, as −***Movement***, with the implication that, like the subject displacement in the English active (i.e., movement from VP-internal position as in (5) above), this operation is irrelevant to the aphasic comprehension deficit. It is to this categorization that our critics object. Taken to task, we now justify our categorization.

How can we discover the derivation of Germanic passive? In English, surface consequences of movement are visible, as passivization moves the object of the active across the verb, (eventually) placing it in subject position. It is not difficult (though non-trivial) to be convinced that the object moves leftwards, thereby crossing the verb and leaving a trace to its right:

(7) *The window*$_2$ [$t'_2$ was **closed** $t_2$] (by him)

In Germanic, however, the effects of movement are less visible—the object of the active (8a) and the subject of the passive (8b) are on the same side of the main verb.

(8) a. Er hat *alle Fenster* **geschlossen**
b. *Alle Fenster* wurden **geschlossen**

Yet reliance on surface order is not enough, especially in Germanic languages. Invisible movements may have taken place during the derivation, which might affect comprehension in aphasia. We need to rule these out. Still, our task here is limited: we do not need to discover how passive in German/Dutch is derived, but rather, to decide whether or not there are movement operations that set it apart from its active counterpart in ways that interact with the comprehension deficit in Broca's aphasia. We claim that no such operations are involved, because unlike English (5), movements in both active and passive in Dutch/German do not cross the verb, leaving both constructions on a par with respect to the aphasic deficit. We use Scope Freezing to demonstrate this parity.

Above we concluded that Scope Freezing tests whether two SBEs crossed each other. It might thus be the right tool for the job of comparing active sentences to their passive counterparts, in order to see whether order-changing operations have taken place at some stage of the derivation. Take an active fragment like (9) that contains a verb and an object which are SBEs but is nonetheless scopally unambiguous due to Scope Freezing (cf. the schematic configuration (4a)). If passivization involves additional (albeit invisible) crossing, scope ambiguity is predicted (as one SBE crosses another, like in (4b)). If, however, passivization does not involve additional crossing, no meaning difference between the active and the passive is expected, hence no ambiguity in the latter. As Wurmbrand (2003) observes, both active and passive in (9) have one meaning only:

(9) a. Weil er *alle Fenster* **vergessen** hat zu schließen...
since he all windows(-ACC) forgotten has to close...
'since he forgot to close all the windows'
b. Weil *alle Fenster zu* schließen **vergessen** wurden...
since all windows (-NOM) to close forgotten were-PL...
'since they forgot to close all the windows'

The Scope Freezing diagnostic, that tests for order-changing movement, shows that passive and active in Germanic are on a par. This brings us to the cross-linguistic difference we sought to underscore: in English, Spanish, and Hebrew, passivization moves the object leftward so that it crosses the verb, and an active/passive performance difference is observed in aphasia; in German/Dutch passive, by comparison, the moved object does not cross the verb in the same manner, no performance difference is observed for the aphasic patients. Thus, while there is movement in German/Dutch passive, the absence of crossing makes this movement irrelevant to the aphasic deficit. For this reason, movement in Dutch and German passive was ignored in our analysis (cf. Grodzinsky, in press, for one possible analysis).

*4.3. Other syntactic issues that regard Germanic languages*

Dutch allows the following alternation in passive sentences, in which the optional *by*-phrase can be placed in two different locations:

(10) a. *Door het meisje* wordt de jongen gekust
'*By the girl* was the boy kissed'
b. De jongen wordt *Door het meisje* gekust

One experiment (Friederici & Graetz, 1987) tested aphasic performance on both configurations, finding no difference—the fronting of the *by*-phrase did not diminish the patients' (otherwise high) level of comprehension performance on this construction. While all the patients we discuss were tested on version (10a), and only a subset of the patients was tested on (10b), in our paper we chose to exemplify Dutch passive by sentence (10b), as it simplified the presentation. In particular, we explicitly suppressed the contrast in (10), which is irrelevant to our quantitative analysis (cf. note 12 in our target paper). With that, Bastiaanse and Zonnenfeld disagree. They claim that the use of the two manifestations of this alternation (i.e., the ±fronting of the *by*-phrase) by Friederici and Graetz disqualifies their study and excludes it from consideration, and that, moreover, the TDH has the wrong prediction in this case. That the experiment is relevant to our discussion almost goes without saying, because it included both types (10a,b) with the same results. We now examine the interaction between the position of the *by*-phrase in Dutch passive and the TDH.

At issue is the manner by which $\theta$-roles reach the object of the by-phrase in passive. The rich literature on this issue (e.g., Baker, Johnson, & Roberts, 1989; Fox & Grodzinsky, 1998; Jaeggli, 1986; Marantz, 1984; to mention just a few) has these conclusions (relevant to our context): a. *By*-phrases in passive are typically assigned the verb's external $\theta$-role; how this happens is not clear, but the assignment process seems to involve the passive morphology on the verb. b. Objects of *by*-phrases have $\theta$-roles even when the predicate is not a verb, hence there is no passive morphology around, from which this $\theta$-role could originate; in these cases, the preposition *by* seems to assign a default role, that is either agent or instrument, never experiencer or source (Fox & Grodzinsky, 1998; Jaeggli, 1986):

| | |
|---|---|
| (11) a. The destruction of the city (by the enemy) | agent |
| b. The destruction of the city (by lightning) | instrument |
| c. The fear of John (*by Bill) | experiencer |
| d. The gift to John (*by Bill) | source |

The preposition *by*, then, is inherently capable of assigning an agent or an instrument role, as evinced by the role of its object in nominals, that is, in the absence of passive morphology.

Consider now fronted *by*-phrases in Dutch passive (10a). It is not clear that they are derived by movement, as they always have a quasi-adjunct status (cf. Grimshaw, 1990) Still, we will assume with Bastiaanse and van Zonnenfeld that they are derived by movement, and hence leave behind a trace that mediates the assignment of a $\theta$-role (from the passive morphology). Under these circumstances, it would follow that in Broca's aphasia, the fronted *by*-phrase is disconnected from its $\theta$-role, as Bastiaanse and van Zonnenfeld observe. Yet, in light of the fact that *by*-phrases can be thematically active in the absence of this $\theta$-role (as we saw in (11)), this disconnection may be inconsequential to the deficit in Broca's aphasia. That is, if the verb at issue is agentive, then even if the fronted *by*-phrase does not receive a $\theta$-role due to trace deletion, the argument inside it would still be agent, as this is the default role that *by* assigns, as we have seen in (11). Thus in the context of agentive verbs, Bastiaanse and van Zonnenfeld's observation leads to a conclusion opposite from theirs, namely that the TDH predicts normal performance, precisely as Friederici and Graetz found.[4]

There is, however, a subtle prediction that can be derived here: Although in the context of passivized agentive verbs, the fronting of Dutch *by*-phrases should not diminish comprehension performance in Broca's aphasia, if a movement operation is involved (and recall, this in itself is not clear), then performance should drop sharply when *by*-phrases of passivized experiencer verbs (of the type *see*, *hear*, *love*, *hate*, *fear*, etc.) are fronted. This would happen because, subsequent to trace deletion, the object of the *by*-phrase should be dissociated from the experiencer role that the predicate assigns, and the preposition *by* would assign it an agent role by default, and a deviant representation that contains agent instead of experiencer would result. In the absence of auxiliary assumptions, then, we would thus expect to obtain differential performance levels on the following, although performance should be normal on both instances when the *by*-phrase would be *in situ* (cf. Fox & Grodzinsky, 1998, for somewhat similar discussion in the context of children's grammar):

[4] The Default Strategy should be considered here as well. It is not clear whether it is invoked, because the object of the *by*-phrase is not thematically lacking (or "dangling," as it has been called), having received the agent role from *by*; yet even if the strategy is invoked, it alone can help obtain the desired result: A fronted *by*-phrase brings its object to sentence-initial position, and the strategy assigns it the agent role. Either way, the position of the *by*-phrase in agentive verbs is orthogonal to the comprehension deficit in Broca's aphasia.

(12) a. *Door het meisje* wordt de jongen gekust
‘*By the girl* was the boy kissed’
b. *Door de meisje* wordt het jongen gevreesd
‘*By the girl* was the boy feared’[5]

In any event (and pending the results of an experiment to that effect), the Friederici and Graetz (1987) study only involved agentive verbs, resulting in high performance levels, as the TDH predicts.

Bastiaanse and van Zonnenfeld raise another issue—Germanic V2. They point out that Dutch and German are known to be SOV languages, and that verbs in main clauses are nonetheless found in the well-known second position of the main clause, leading to an analysis according to which in main clauses, verbs invariably move to second position (S O V $\rightarrow$ S $V_i$ O $t_i$). As this movement leaves a trace behind, our critics argue that the TDH should lead to problems in simple active declaratives in German and Dutch, contrary to fact.

This argument would be potentially interesting had the TDH claimed that traces of verbs are deleted. However, the empirical evidence speaks against that: In various studies, Broca’s aphasics have repeatedly demonstrated sensitivity to verb movement (or more generally, to head movement, cf. Grodzinsky & Finkel, 1998; Linebarger, Schwartz, & Saffran, 1983; Lonzi & Luzzatti, 1993; see Grodzinsky, 1995a, 1995b for discussion of the restrictive nature of trace deletion; see also Grodzinsky, 2004 for discussion of the TDH in the context of German main clauses). This differential sensitivity is not surprising, given the different theoretical status of phrasal traces and those of head movement (cf. Chomsky, 1995).

### *4.4. Experimental issues regarding the Germanic passive*

The two preceding arguments regarding Germanic (fronted *by*-phrases, main verbs in second position) lead Bastiaanse and van Zonnenfeld to reject both Friederici and Graetz’s and Kolk and van Grunsven’s (1985) experiments from consideration. As we have seen, neither argument justifies the exclusion of these from the pool. Bastiaanse and van Zonnenfeld then discuss a new study (Bastiaanse and Edwards, 2004) of Dutch active/passive, that putatively failed to replicate past results to which we alluded above. Reanalyzing a subset of the new results, we compared SVO actives to passives.[6] Group results ($n = 13$): $\mu_{active} = 0.83$, $\sigma_{active} = 0.21$; $\mu_{passive} = 0.71$, $\sigma_{passive} = 0.239$; the probability of a difference between the two on our hypothesis testing, $p = .4$. These results agree with our conclusion that Dutch actives and passives are not different in Broca’s aphasia. Compare with our results for English ($n = 27$; $\mu_{active} = 0.829527$, $\sigma_{active} = 0.133529$; $\mu_{passive} = 0.630427$, $\sigma_{passive} = 0.205748$; $p = .0014$). This contrast is evident in the fitted $\beta$-curves (active—perforated red line, passive—solid black line) (see Fig. 2).

The results, then, are clear: the active/passive difference observed for English is not found in the Dutch/German case. The means and variances for actives are similar, and hence the cross-linguistics effect is due to passive—low in English and higher for German/Dutch. Remarkably, the difference between English and German/Dutch scope ambiguities correlates with the aphasic performance.[7] And while we do not have an account, the correlation itself seems worth exploring.

## 5. Italian active/passive: A puzzle

We now move on to another potentially interesting cross-linguistic difference, from a study by Caramazza et al. (2005). They present a data set culled from comprehension scores of 38 Italian Broca’s aphasics. While error rates on passive were higher than those for active in this data set, no significant difference was found between the two conditions.

Caramazza et al. used statistical tests that are based on paired individual performances (mostly $\chi^2$). For the reasons we gave above, a broader picture is necessary, and we thus reanalyze these data with our analytic tools. Our analysis indeed confirms that at the group level, there is no significant difference between performance on active and on passive (group ($n = 38$) results: $\mu_{active} = 0.77$, $\sigma_{active} = 0.13775$; $\mu_{passive} = 0.69$, $\sigma_{passive} = 0.167033$; the probability for a difference between the two on the $\beta$-distribution-based hypothesis testing, $p = .16704$). A similar result also seems to hold for Broca’s aphasics’ performance in another recent study in Italian which featured the active/passive contrast (Luzzatti et al., 2000, group ($n = 11$) results: $\mu_{active} = 0.802449$, $\sigma_{active} = 0.178722$; $\mu_{passive} = 0.652488$, $\sigma_{passive} = 0.259438$, $p = .319543$). These results, however, contrast with those obtained for our English data set ($n = 27$), already reported above. The contrast is evident in the fitted $\beta$-curves (active—perforated red line, passive—solid black line) (see Fig. 3).

[5] We thank Ineke van der Meulen for her help with the Dutch examples. Christiane Ulbrich informs us that the same is true for German:

(i) *Von einem* Mädchen wurde John geküsst.
By a girl John was kissed.
(ii) *Von einem* Mädchen wurde John geliebt.
By a girl John was loved.

[6] See Bastiaanse and Edwards, (2004, p. 101, Table A1 in the Appendix). We compared the proportion of correct responses (the “+” column) in their AVT and TAV conditions.

[7] A similar (though not unproblematic) result has been obtained by a study of English and German speaking children. Aschermann, Gülzow, and Wendt (2004) compared patterns of acquisition of active and passive in the two languages, and found that the latter children attain passive earlier than the former.

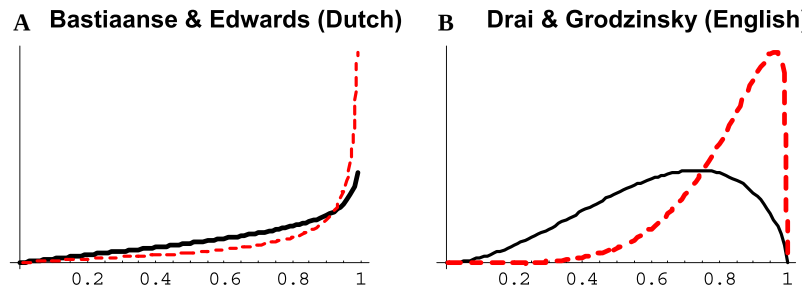


Fig. 2. $\beta$-Distributions for the recent active/passive Dutch study and for English.

Like the Dutch/German case, no significant active/passive difference was found. Our analysis enables us to look at group means, and discover a difference between Dutch/German and Italian: Unlike the Dutch/German case, the absence of an effect is not attributable to passive. Rather, the mean Italian performance scores for *active* sentences are somewhat lower than English, whereas those for passive are somewhat higher.

What could be the reason(s) for the remarkable difference between English and Italian? Caramazza et al. do not answer this question, so we can at least consider the logical possibilities, as we did in the Dutch/German case: (i) the Italian and English tests differ in a relevant way; or (ii) the languages differ in a way that interacts with the patients' deficit; or (iii) the patient groups are different.

No obvious solution is in sight: First, Caramazza et al. provide one example of a stimulus, but not a complete list, so we are unable to comment on the test. Second, at this point we are unaware of anything that sets the Italian passive apart from its English counterpart in any relevant respect. Lastly, there might be something special about the particular group of patients that were tested, yet neither this paper, nor Luzzatti et al.'s, contain a suggestion or a hint in this direction. The mystery of Italian passive thus awaits solution. Importantly, and contrary to Caramazza et al.'s suggestion, the data are highly structured, setting languages apart in ways that lead to the formulation of new research questions. The statistical tools we offer, and theoretical tools such as the TDH, seem indispensable when such problems are contemplated. Moreover, only a group-level analysis provides a vantage point from which the present result can be seen: group means and variances can be computed, and a comparison between the two Italian groups of patients can be made, which underscores a pressing need to understand the Italian case.

## 6. Anatomical variation

Amunts and Willmes note that lesion localization in aphasia is carried out through neuroimaging instruments, whose current resolution only enables the identification of topographic landmarks, but not cytoarchitectonic borders. Functional compartmentalization, they note, is more likely to align with the latter. This limitation, Amunts and Willmes conclude rather gloomily, diminishes the likelihood that lesion analysis in aphasia would scope Broca's region precisely, hence we still have a way to go before we can use aphasia to establish precise structure/function parallels. Ideally, then, having access to cytoarchitectonic borders at the individual subject (or patient) level would improve structure/function mapping.

This perspective is quite plausible. In fact, as the commentary shows, anatomical work by Amunts and her colleagues (e.g., Amunts et al., 1999; Amunts & Zilles, in press) has improved our understanding of the anatomical variability that exists in the language regions. Yet does it evacuate current mapping methods of their content? We would argue that our results suggest that Amunts and Willmes's bleak perspective is not in place, and that a cautiously optimistic outlook is the right one.

Before attempting to adjudicate between gloom and optimism, however, we must first backtrack a bit, be-

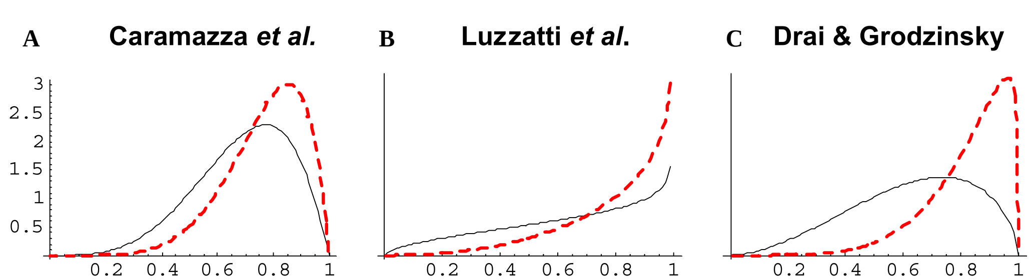


Fig. 3. $\beta$-Distributions for the active/passive contrast in two Italian studies and in English.

cause to begin with, the diagnosis of patients in our data set was determined on mostly *behavioral*, not anatomical, grounds. We must, therefore, consider the issue of anatomy/function alignment in two steps: First, we must see how closely clinical diagnosis is related to lesion site (if at all); second, we must see what consequences there are to Amunts and Willmes' observation that the alignment between macro-anatomically defined lesion boundaries and microscopically delineated cytoarchitectonic borders is poor.

Clinical signs and anatomy first. Diagnostic tests of aphasia (typically the BADE, WAB, AAT and the like) are not precise tools. Yet they do give a reasonably clear idea regarding the involvement of certain brain regions. From the perspective of gross anatomy, the area implicated in Broca's aphasia includes Broca's area, but goes beyond it. Broca's aphasia is taken to ''encompass most of the operculum, insula, and subjacent white matter, exceeding Broca's area'' (Mohr, 1976, p. 202). And although there are questions as to the precise borders of this area, and their relation to severity of non-fluent aphasia (see Naeser, Palumbo, Helm-Estabrooks, Stiassny-Eder, & Albert, 1989), the area described by Mohr seems necessary for producing the classical picture of Broca's aphasia consistently. Thus, while Naeser et al. find that frontal patients with no speech at all, or only with stereotypes, are anatomically distinct from classical Broca's patients in that they suffer not only from damage to Broca's area, but also to subcortical structures and connections (around the peri-ventricular white matter), there does not seem to be a serious challenge to the close connection established between the signs of Broca's aphasia—as picked up by clinical tests—and Broca's area and its vicinity.

As Amunts points out, the size of the area described by Mohr might be explained by the cytoarchitectonic variation her work has discovered. It is quite likely that Mohr's (and others') conclusions regarding the extent of the anterior language areas are a consequence not of the overall size of the area that produces Broca's aphasia, but rather, of individual variation in the positioning of the relevant neural tissue. Be it as it may, the proof (or at least a flavor thereof) is in the pudding: our reliance on clinical diagnostic tests (and imaging information when available) helps us focus on the left anterior frontal regions, imprecise as it might currently be, by still leading to a behaviorally stable and linguistically coherent picture. Thus, while precise cytoarchitectonic mapping of syntactic ability is a yet unaccomplished desideratum (as is a cytoarchitectonic map of motor, visual, or object recognition ability), we are perhaps closer to it than are many other areas in cognition: the structure found in behavioral lesion data, together with precise anatomical discoveries like those due to Amunts and her colleagues, brings us closer to such maps. Once enhanced instrumentation would give us access to cytoarchitectonic information in vivo, we will hopefully be able to see through anatomical variation, just as we are now beginning to abstract away from behavioral variation through statistical and linguistic tools.

## 7. Alternative linguistic analyses

Two alternative analyses of the data were proposed, both aimed at deriving production and comprehension patterns: Bastiaanse and van Zonnenfeld propose the Derived Order Problem Hypothesis (DOP-H), and Friedmann attempts an extension of the Tree-Pruning Hypothesis (TPH) to comprehension. Seeking to keep the focus of this paper on the quantitative analysis of comprehension in Broca's aphasia, we will touch on each proposal briefly.

The DOP-H claims that Broca's aphasic patients have a problem with ''derived word orders.'' As far as one can tell, this proposal is for the most part indistinguishable from Schwartz, Linebarger, Saffran, and Pate's (1987) ''mapping hypothesis,'' the claim that the comprehension of constructions that contain moved constituents—those to which the mapping of $\theta$-roles onto arguments is indirect (mediated by a trace)—is poor in Broca's aphasia. This theory, built around syntactic movement, is a rather loose variant of the TDH, because at best, it partitions the data into −***Movement*** constructions that are ''well'' comprehended and +***Movement*** constructions that are comprehended ''poorly.'' Crucially, it cannot derive actual performance levels (cf. Grodzinsky, 1990, p. 70, 2000 for critique along these lines). From a current perspective, the Mapping Hypothesis is empirically inadequate even in its own terms, because a broad range of comprehension results have been added to the literature since it was proposed, calling for a more refined approach. Among these are the comprehension contrasts documented for quantified vs. non-quantified subjects in passive, both of which are +***Movement*** (Balogh & Grodzinsky, 2000; Saddy, 1995); similarly, it is incapable of handling the performance asymmetry regarding *which-N* vs. *who* object questions (both of which are +***Movement***, see Hickok & Avrutin, 1995); and finally, it fails on the English vs. Dutch/German differences in passive that were discussed above.

Interestingly, there are some relatively new results that do distinguish the DOP-H from the Mapping Hypothesis. In particular, as the notion ''derived'' seems germane to the DOP-H, there is no reason to expect ''good'' performance on any *active* sentence, because these constructions are ''derived,'' that is +***Movement***, as in general, subjects are thought to move to their ''surface'' position from a lower position, internal to the VP (cf. Koopman & Sportiche, 1991, among many others). This prediction is of course contrary to fact. Likewise, patients' inability to detect violations of constraints on movement predicts failure in detecting violations of constraints not only on phrasal movement, but also on head movement. Massive evidence to the contrary exists in the literature (Grodzinsky & Finkel, 1998; Linebarger et al., 1983; Lonzi & Luzzatti, 1993).[8] The interpretation we offer of the DOP-H, then, seems to have little empirical support, if any.

Friedmann also tries to derive one aspect of the results from a deficit model designed for production patterns—the TPH. Based on several studies she reviews, she concludes that Broca's aphasic patients who were tested on both object relatives and passives were either at chance in both structures, or only on object relatives, yet, crucially, never at chance on passive but not object relative.[9] She then proposes to account for this pattern by appealing to the severity aspect of the TPH. This account, originally designed for production, assumes a particular ordering of functional categories in a phrase marker, and claims that in Broca's aphasia, the top part of the tree is "pruned"; further, it seeks to identify degrees of severity in production with the relative height of the locus of impairment. The higher the locus of impairment, the lesser the degree of severity. More precisely: "For $P_1, P_2, \ldots, P_n$, different variants of the syndrome, $P_i$ is *more severe than* $P_{i+1}$ iff $N_i$, the node impaired in $P_i$, is contained in the c-command domain of $N_{i+1}$, the node impaired in $P_{i+1}$" (Friedmann & Grodzinsky, 1997, p. 421). She proposes to extend the TPH to comprehension thus: In passive, the subject moves to the specifier position of IP; in object relative clauses, the object trace is bound by an operator in CP (which dominates IP); finally, IP dominates T. A T impairment affects all parts of the tree above it, including IP and CP, which would result in problems in passives and object relatives. A milder impairment—one that only affects nodes that c-command TP—might leave the subject position intact, hence capable of housing the subject of the passive. This milder impairment would still have detrimental consequences to object relatives, because the CP layer would still be impaired, which would hamper the trace binding operator in CP from fulfilling its function. This extension to comprehension in Broca's aphasia, Friedmann proposes, would result in an elegantly symmetric account of comprehension and production that links the degrees of severity in both modalities.

This idea is very interesting, but it leaves many questions unanswered. First, consider an alternative that must be ruled out. In the target article we separated the analysis of mono-clausal and bi-clausal sentences, reasoning that aphasics might generally have more difficulty in analyzing the latter. If so, then the putative difference between passive and object relatives does not stem from pruned trees, but rather, from something more generic. And while there are numerical issues that must be resolved (e.g., what exactly is the contribution of difficulty to overall performance? If it exists, how is it to be measured, and factored out of the linguistic deficit?),[10] it seems that this possibility must be ruled out before a TPH-based account could be considered.

Second, consider two puzzles that emerge once the TPH extension is fleshed out. a. If the comprehension deficit has an incremental aspect to it—going higher and higher in the tree as severity goes down, then why does it "jump" nodes? Friedmann adopts a conception of the phrase marker in which operators are above subjects, which in turn are above the position that represents tense. Yet the deficit she proposes "jumps" nodes, being located either at T, or at C. An elegant incremental aspect of the TDH would be severely diminished under such a conception, b. In what way is the operator that binds the object trace impaired, so as to bring about chance performance? Since Broca's aphasics have virtually no comprehension problems in subject-gap relative clauses (and clefts), what is it about the impairment of this operator (subsequent to C-pruning) that singles out the object-gap structures, and depresses comprehension performance in them? Finally, as the relative head c-commands the relative's impaired CP, and given that this same head is also a subject or an object of a main clause, is the comprehension of these main clause impaired in Broca's aphasia, even in subject-relatives?

Furthermore, there are comprehension asymmetries that cannot be handled by a mere appeal to syntactic configuration (as an unamended TPH would have it), which makes no distinctions among category types and movement types. Examples such as the English vs. Germanic passive, and the Hickok/Avrutin *which-N*

[8] Bastiaanse and van Zonnefeld attempt to restrict their account by stipulating that verb movement does not affect comprehension since it "does not affect the meaning of the sentence." Perhaps, yet several questions remain: first, why would change in meaning be a criterion for relevance of the DOP-H? Second, if so, why would such a restriction on the application of the DOP-H be invoked in reception tasks that do not require interpretation? That is, why do Broca's aphasics detect violations of certain constraints on verb and other head movement, but not XP-Movement? Third, perhaps most poignantly, if change in meaning is required, then why do Broca's aphasics have trouble in English passive, but not in its active counterpart which presumably has the same truth-conditions?

[9] An examination of our data base in light of this proposal is beyond the scope of this paper, but the reader is invited to do so.

[10] This is yet another type of issue which can be resolved through calculation: That is, an account that banks on the difficulty of bi-clausal sentences would predict an overall decrease in the total proportion correct of *both* subject-gap and object-gap relatives when compared to (at least non-scrambled actives) and (moved) passives. The interested reader is once again invited to do these on our data set.

vs. *who* questions come immediately to mind. The TPH might be modified to accommodate such data, perhaps by an appeal to differences in the semantic content of different functional categories (see Valeonti, Economou, Kakavoulia, Protopapas, & Varlokosta, 2004, for a recent proposal along these lines). These questions may all have answers, and in this respect Friedmann's interesting proposal has the potential of opening new venues.

## 8. Concluding remarks

The critical commentary on our work was most helpful: Amunts and Willmes, Toraldo and Luzzatti and Caplan et al. sharpened issues that pertain to our quantitative analysis, while Amunts and Willmes pointed out anatomical limitations of our perspective; de Bleser, Burchert, and Schwarz and Bastiaanse and van Zonnenfeld forced us to be more explicit about our assumptions regarding the Germanic passive, whereas the latter added new data on Dutch to the pool; Caramazza et al. added more data from Italian passive. Finally, Bastiaanse and van Zonnenfeld and Friedmann considered alternative accounts. None of the concerns of our critics seem to have harmed our method, or the account it is designed to support, and in fact, left our claims for the most part unchallenged. Moreover, these contributions helped us increase our 69-patient data base with scores of 38 patients from Caramazza et al., 11 patients from Luzzatti et al., and 13 patients from Bastiaanse and Edwards' studies into the pool of raw data, which further strengthened our approach. And while it seems that our analysis still stands, we are pleased to see that our methods, when applied to new data, have added new, intriguingly complicated facts to the (already complex) picture. Arguably, while the (poorly understood) passive construction may have received too much attention in this debate, the cross-linguistic investigation did expose new facts. Hopefully, the variability debate will stimulate researchers to ask new questions, use new linguistic material, and broaden our theoretical and empirical horizons in novel ways.

## References


Aschermann, E., Güzlow, I., & Wendt, D. (2004). Differences in the comprehension of passive voice in German and English speaking children. *Swiss Journal of Psychology, 63*, 235–245.

Avrutin, S. (2001). Linguistics and agrammatism. *Glot International, 5.3*, 1–5.

Baker, M., Johnson, K., & Roberts, I. (1989). Passive arguments raised. *Linguistic Inquiry, 20*, 219–251.

Balogh, J., & Grodzinsky, Y. (2000). Levels of linguistic representation in broca's aphasia: Implicitness and referentiality of arguments. In R. Bastiaanse & Y. Grodzinsky (Eds.), *Grammatical disorders in aphasia: A neurolinguistic perspective*. London: Whurr Publishers.

Caramazza, A., Capasso, R., Capitani, E., & Miceli G. (in press). Patterns of comprehension performance in agrammatic Broca's aphasia: A test of the Trace Deletion Hypothesis. *Brain and Language*.

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

Fox, D. (2003). On logical form. In R. Hendricks (Ed.), *Minimalist syntax*. Oxford: Blackwell.

Fox, D., & Grodzinsky, Y. (1998). Children's passive: A view from the *by*-phrase. *Linguistic Inquiry, 29*, 311–332.

Friederici, A., & Graetz, P. (1987). Processing passive sentences in Aphasia: Deficits and strategies. *Brain and Language, 30*, 93–105.

Friedmann, N., & Grodzinsky, Y. (1997). Tense and Agreement in agrammatic production: Pruning the syntactic tree. *Brain and Language, 56*, 397–425.

Frey, W. (1993). *Syntaktische Bedingungen für die semantische Interpretation: Uber Bindung, implizite Argumente und Skopus*. Berlin, Germany: Akademie.

Grodzinsky, Y. (1984). Language deficits and linguistic theory. Doctoral dissertation, Brandeis University.

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

Grodzinsky, Y. (1995a). A restrictive theory of trace deletion in agrammatism. *Brain and Language, 51*, 26–51.

Grodzinsky, Y. (1995b). Trace-deletion, $\theta$-roles, and cognitive strategies. *Brain and Language, 51*, 469–497.

Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences, 231*, 1–71.

Grodzinsky, Y. (2004). Variation, canonicity and movement in aphasic comprehension. Paper presented at the EURESCO "Science of Aphasia V" conference, Potsdam, September. (Handout available on http://freud.tau.ac.il/~yosef1/).

Grodzinsky, Y. (in press). A blueprint for a brain map of syntax. In Y. Grodzinsky & K. Amunts (Eds.), *Broca's region*. New York: Oxford University Press.

Grodzinsky, Y., & Finkel, L. (1998). The neurology of empty categories: Aphasics' failure to detect ungrammaticality. *Journal of Cognitive Neuroscience, 10.2*, 281–292.

Hickok, G., & Avrutin, S. (1995). Comprehension of Wh-questions by two agrammatic Broca's aphasics. *Brain and Language, 51*, 10–26.

Jaeggli, O. (1986). Passive. *Linguistic Inquiry, 17*, 587–622.

Kolk, H., & van Grunsven, M. (1985). Agrammatism as a variable phenomenon. *Cognitive Neuropsychology, 2*, 347–384.

Koopman, H., & Sportiche, D. (1991). The position of subjects. *Lingua, 85*, 211–258.

Krifka, M. (1998). Scope-inversion under the rise-fall contour in German. *Linguistic Inquiry, 29*, 75–112.

Linebarger, M., Schwartz, M., & Saffran, E. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13*, 361–393.

Lonzi, L., & Luzzatti, C. (1993). Relevance of adverb distribution for the analysis of sentence representation in agrammatic patients. *Brain and Language, 45*, 306–317.

Marantz, A. (1984). *On the nature of grammatical relations*. Cambridge, MA: MIT Press.

Mohr, J. P. (1976). Broca's area and Broca'a aphasia. In H. Whitaker & H. A. Whitaker (Eds.). *Studies in neurolinguistics* (Vol. 1, pp. 201–235). .

Naeser, M. A., Palumbo, C., Helm-Estabrooks, N., Stiassny-Eder, D., & Albert, M. L. (1989). Role of the medical subcallosal fasciculus and other white matter pathways in recovery of spontaneous speech. *Brain, 112*, 1–38.

Saddy, D. (1995). Variables and events in the syntax of Agrammatic speech. *Brain and Language, 50*, 135–150.

Sauerland, U. (2003). On quantifier raising in German. Ms. Universität Tübingen.

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

Valeonti, N., Economou, A., Kakavoulia, M., Protopapas, A., & Varlokosta, S. (2004). The breakdown of functional categories in Greek aphasia. Poster presented at *Science of Aphasia V*, Potsdam.

Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics, 31.4*, 949–952.

Williams, D. A. (1988). Estimation bias using the beta-binomial distribution in teratology. *Biometrics, 44*, 305–309.

Wurmbrand, S. (2003). A-Movement to the point of no return. In *Proceedings of NELS 33*. Amherst, MA: Umass GLSA.