Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

Brain and Language 96 (2006) 117–128

Brain and Language

www.elsevier.com/locate/b&l

Target article

A new empirical angle on the variability debate: Quantitative neurosyntactic analyses of a large data set from Broca's Aphasia ☆

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

[a] Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel
[b] McGill University, Department of Linguistics, 1085 Dr. Penfield Avenue, Montreal, Que., Canada PQ H3A 1A7
[c] Department of Psychology, Tel Aviv University, Tel Aviv 69978, Israel

Accepted 26 October 2004
Available online 22 August 2005

Abstract

Behavioral variation in Broca's aphasia has been characterized as boundless, calling into question the validity of the syndrome-based schema and related diagnostic methods of acquired language disorders. More generally, this putative variability has cast serious doubts on the feasibility of localizing linguistic operations in cortex. We present a new approach to the quantitative analysis of deficient linguistic performance, and apply it to a large data set, constructed from the published literature: Comprehension data of 69 carefully selected Broca's aphasic patients, tested on nearly 6000 stimulus sentences, were partitioned in different ways, and subjected to a series of analyses. While a certain amount of variability is indeed evident in the data, our quantitative analyses reveal a highly robust selective impairment pattern for the group: the patients' ability to analyze syntactic movement is severely compromised, in line with the Trace-Deletion Hypothesis. Further analyses suggest that patients' performance on no-movement sentence types exhibits less variation than on sentences that contain movement. We discuss the clinical and theoretical implications of our results.

Keywords: Aphasia; Variability; Syntax; Quantitative analysis; Confidence interval; β-Distribution

---

1. The problem

For over a century, the question of patient grouping has occupied the best minds in behavioral neurology. In many, if not most instances, it took the shape of a debate over the typology of aphasic syndromes, and whether or not these are truly distinguishable from one another. As Mohr's well-known review of Broca's aphasia (1978) notes, this syndrome has survived a host of classificatory attempts, from Dejerine's to Luria's. Debates regarding diagnosis, localization, and classification have long been with us, and do not seem to end. In recent past, many researchers have focused on putative similarities and differences between aspects of Broca's and Wernicke's aphasia (i.e., whether "agrammatism" in the speech production of the former syndrome is a deficit different from "paragrammatism" that the latter is said to exhibit, cf. Kolk & Heeschen, 1992, and whether comprehension patterns in the two groups are truly distinct, cf. Kolk, 2004; Zurif, 1995, for opposing views). Yet the debate has also touched the question of uniformity within a syndrome, lately narrowing down to one important issue: is the observed behavior of patients accorded a diagnostic given label not only distinct from others,

---

☆ Supported in part by Canada Research Chairs, Canada Foundation for Innovation, and a McGill VP-Research internal grant. D. Drai'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.

[*] Corresponding author. Fax: +1 514 398 7088.
E-mail address: Yosef.Grodzinsky@mcgill.ca (Y. Grodzinsky).

0093-934X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2004.10.016

but also, sufficiently similar so as to justify our grouping them together through the traditional syndrome schema? Discussion, then, has shifted: from inter-syndromic differences, it has moved to examine intra-syndromic similarities, with Broca's aphasia being the main case in point. And here, positions differ: some researchers (e.g., Avrutin, 2001; Goodglass, 1993; Grodzinsky, 2000; Zurif, 1995) view Broca's aphasia not only as a distinct clinical entity, but also, as an important tool for discerning the functional linguistic role of Broca's area; others (e.g., Berndt, Mitchum, & Haedinges, 1996; Caramazza, 1986; Caramazza, Capitani, Rey, & Berndt, 2000) have denied the validity of the traditional syndrome schema, arguing that no structure can be discerned in the performance of these patients, when taken as a group. To bolster their position, the latter researchers have documented performance instability, purportedly showing how a fixed clinical diagnosis leads to seemingly boundless behavioral variation across patients (in a relevant behavioral domain).

Both groups have claimed that their position is empirically motivated and based on data that exist in the experimental literature. It is on this puzzle that the present paper is focused. Since at most, one of the two views in the variability debate is correct, it is extremely important to determine what the empirical record really is, especially in light of the variation-oriented view: if true, it speaks against generalizations over these data, and the validity of traditional diagnostic methods. It thus undermines both the clinical schema, and the feasibility of localizing linguistic operations in cortex through lesion data.

Below, we employ novel quantitative methods to a large database in order to find structure in the comprehension performance of patients classified as Broca's aphasics. Our database was compiled from carefully selected, published test scores of patients with identical diagnosis of Broca's aphasia. As the variability debate is concerned with the correct interpretation of a data set, it is crucial to characterize the data set and its limitations, and also to define precisely what counts as an empirical validation of a thesis about the presence (or absence) of structure in the empirical record. We therefore outline the general logic of our analysis before embarking upon a detailed description of our procedure.

Our starting point is that performance variation within a group of patients in itself does not preclude the existence of structure in their deficit. Thus in aphasia the data may present inter-patient variability, but the challenge for us is to try and discover commonalities at the group level in the face of this variability. That is, we ask whether group analyses can discover different patterns of performance on different tasks. For that, we constructed a database from raw performance scores of Broca's aphasics—the best-studied aphasic syndrome, hence at the heart of the variability debate. The experiments we discuss measure error rates in forced binary choice comprehension tasks that involve a variety of sentence types. The observed performance level of a particular individual on a given type of sentence can thus be seen to indicate the probability that this patient succeeds. Each patient is identified by a probability parameter for every sentence type. Thus per each sentence type, our group of Broca's aphasic patients has a distribution of these probabilities. Since we have several types of sentences, we have several such distributions for the group, which we can now analyze. We devised a rigorous statistical test to detect differences between such distributions. In other words, we explicitly study the distributions of probabilities of successful performance and seek to infer significant differences of group performances on different types of sentences. This, we believe, dispels most of the obscurity as to the structured nature of the observed performances: if a comparison of performances on two types of sentences yields a significant difference in the corresponding distributions of probabilities we conclude that Broca aphasics are differentially impaired with respect to this contrast.

Before proceeding, let us emphasize that ours is not a meta-analysis: we devised a method for the analysis of raw scores, and this is the nature of our database. Furthermore, our approach does not contradict any previous finding: while claims regarding variation in the empirical record do exist, no one has carried out a large-scale quantitative analysis such as ours. And while our approach does admit the existence of variability (indeed it models it), it asks whether differences between the population's distributions of probabilities of success can be assessed despite the variation. This is quite a standard approach in biology, where no one expects clean dichotomies to appear as a result of a manipulation. Our analysis thus inheres in a standard procedure: we ask whether two distributions can be considered equal in view of the results of a sampling. We are not contradicting the results of previous studies through use of a bigger sample; rather, we are enabling—and carrying out—a principled inquiry into the existence of a hitherto controversial contrast (see Drai, in press).

To recapitulate: we have compiled a large data set of raw data of performance scores, and adopted a statistical methodology which gives a precise quantitative meaning to the question of the existence of a significant difference in performance—analyzed at the group level—between two types of sentences. Note that the approach is general and can be applied to any subject group, and to any categorization of the sentences. We did apply our method to performances of Broca's aphasics, for whom we have different existing categorizations of sentences.

We can now forecast our findings:

(1) Despite performance variability within the group, we can define an explicit stochastic model for the group distribution of probabilities (of success on a given type of sentence). Moreover, this characterization enables a rigorous statistical test of the difference (or lack thereof) between pairs of sentence types in the performance of the group.
(2) In our data, this method reveals a highly significant performance difference when the data are categorized by a syntactic principle (whether or not sentences contain a Transformational Movement relation); more traditional ones (Mood [active/passive], Sentence Complexity [high/low]) yield no contrast.

2. Data selection

Aphasiological studies are typically conducted on samples of small size (rarely exceeding $n = 10$). To give our analysis a serious quantitative character we constructed a large data set of sentence level comprehension scores of Broca's aphasics. Furthermore, the analytic methods we developed for the evaluation of raw behavioral test results from multiple experiments presuppose a high degree of homogeneity among studies. To ensure homogeneity in our database, we limited the scope of our investigation to a single type of result from one syndrome: scores obtained from Broca's aphasics in sentence-to-scenario matching experiments that measure error rates in sentence comprehension. While task and clinical criteria were kept constant, sentence stimuli varied in systematic ways that are at the heart of our analysis.

We used clinical diagnosis for patient selection. Did this selection bias our results? While logically possible, we do not see a reason to suppose that this was the case. Our main demonstration below is that one partition method discovers structure in the data, whereas other methods do not. No clinical diagnostic test we use is geared toward such distinctions. These tests are designed to evaluate general communicative skills, and thus focus mostly on speech production, whereas comprehension abilities are evaluated in a very general way. It thus seems that the way clinical tests select patients is orthogonal to our demonstration.

Diagnosis was done on (mostly) behavioral grounds (typically the BDAE, WAB, AAT, and the like), and we thus restrict ourselves to the analysis of behavior. We now move to describe how our database was constructed: scores were selected from the published literature (cf. Appendix A for a list of studies) by a sieve of five highly restrictive criteria. A patient's score was admitted into the study only if all of the following requirements were met:[1]

(a) Positive diagnosis of Broca's aphasia by a well-established test,[2] and sound neurological considerations (i.e., imaging when available, clinical workup otherwise);
(b) Tests investigate comprehension at the sentence level, through a forced binary-choice task ('who did what to whom') with multiple 'semantically reversible' sentences, and with scenario pairs that depict thematic reversals <"*a* does X to *b*," "*b* does X to *a*">;
(c) Detailed descriptions of experimental conditions and procedures are available;
(d) Raw individual patient scores are available;
(e) Identifying information is available (to avoid multiple counting of patients who participated in more than one study).

This rigorous sieve produced a uniform data matrix in terms of patient selection and data acquisition procedures.[3] It contained raw scores from 21 studies, with 233 data points (each being a single patient's raw score on a particular sentence type), which were constructed from 5934 trial sentences of 14 types. Tests were conducted in different locations and times, with 69 aphasic speakers of seven languages (Dutch, English, German, Hebrew, Japanese, Korean, and Spanish). Experiments differed only in number of trials (not always equal per contrast within, let alone between, experiments), and number of successes per experimental condition.[4]

3. A three-way data partition

The data set contains comprehension test scores from two main classes of sentence stimuli: bi-clausal relatives, and mono-clausal actives/passives. The systematic richness thus obtained enabled quantitative comparisons (always among sets of minimal pairs) between three data categorization methods:

(A) The psycholinguistic measure of perceptual ***Complexity*** (HIGH/LOW): a behavioral property of a sentence that seems to correlate with one or more syntactic properties. Widely viewed as a central factor in processing, complexity is thought to rely

---

[1] The complete data matrix may be downloaded from http://freud.tau.ac.il/~yosef1.

[2] E.g. Boston Diagnostic Aphasia Examination, Western Aphasia Battery, Aachen Aphasia Test, etc.

[3] We focused on behavior because it is the element under fire in recent years. Neuroanatomically, we remained at the level of clinical diagnostic tests and structural MR images, that have a reasonably good localizing value.

[4] This dispenses with the need for the use of a weighted effect size based meta-analysis.

on Broca's region as its anatomical substrate (Caplan & Waters, 1999; Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Röder, Stock, Neville, Bien, & Rösler, 2001). Bi-clausal sentences in our data matrix were (high complexity) Center Embedded, and (low complexity) Right-Branching, relative clauses.

(B) The traditional grammar distinction ***Mood*** (ACTIVE/PASSIVE): arguably, brain damaged individuals experience difficulty in processing bi-clausal sentences in general. If so, then contrast (A) may lead to perturbations. Luckily, the best studied pair in aphasiology (active/passive) contains mono-clausal declarative sentences, and may provide a more distilled contrast—***Mood***. In (agentive) active sentences, the Subject is agent-of-action, and passive morphology is absent; in passive, the Subject's being recipient-of-action correlates with morphological marking of passive voice—English-*en* or its analogs. Time and again has the passive construction featured in investigations into the nature of Broca's aphasia (Goodglass, 1968; Grodzinsky, 1984a, 1984b, 1986; Schwartz, Saffran, & Marin, 1980; and many, many others). Early results come almost exclusively from English, a language in which passive tends to correlate with "deviation from canonical ordering of major constituents"—a notion often thought to be equivalent to syntactic movement. More recent studies have obtained cross-linguistic data, from language that enable to tease the active/passive distinction apart from word order changes. We thus chose ***Mood*** as a dimension along which to analyze the data we compiled.

(C) The linguistic notion ***Movement*** (±): simply put, ***Movement*** involves intra-sentential "action at a distance," which cuts the data orthogonally to each of the two contrasts above. Consider the active/passive distinction: in the active sentence '*the horse kicked the rider*,' the verb *kick* determines the semantic roles of the argument immediately preceding it ('*the horse*'), and of the one immediately following it ('*the rider*').[5] In the corresponding passive '*the rider was kicked*' ▶ '*by the horse*,' however, the elements <*kick*, *the rider*> are non-adjacent, and their sequential order is reversed. Still, semantic roles are preserved under this major change—'*the rider*' remains recipient-of-action. The verb '*kick*' assigns a patient role rightwards just like in the active, to the position marked by '▶' (cf. e.g., Marantz, 1984). This means that phonetically, '*the rider*' in passive is phonetically present sentence-initially, but its semantic role is downstream, in '▶'. The two positions <'*the rider*', '▶'> must be related, to ensure correct interpretation during comprehension. This relation is captured by a Transformational Movement rule. Similar considerations hold when two types of relative clauses are compared: '*the nurse helped the horse that kicked the rider*' vs. '*the nurse helped the rider that the horse kicked* ▶.' Some aphasia and fMRI studies have linked ***Movement*** to Broca's region (e.g., Ben-Shachar, Hendler, Kahn, Ben-Bashat, & Grodzinsky, 2003; Grodzinsky, 2000; Zurif, 1995).

Having established these three categorization methods, we proceeded to study them quantitatively. First, we categorized the bi-clausal relative clause subset of the data by ***Complexity***, and then by ***Movement***; second, we categorized the mono-clausal sentences by ***Mood***, and then by ***Movement***. Doing so enabled us to test the relative merit of the different data partitioning methods.

4. Data analysis: Synopsis of methods

The raw data take the following shape: for a given patient, performance on a particular sentence type is given as the couple (number of trials, number of successes). As each patient has several data points, scores can be categorized to obtain different contrasts (A or B or C above). Our analysis first categorizes the raw data in one way, analyzes them quantitatively at the individual and the group levels, and then reanalyzes precisely the same data set after re-categorization. This enables us to carry out numerical comparisons between different partitioning of the data, to determine the categorization of choice—the one in which structure is best discerned. Our analysis proceeds in two rounds (see Appendix B for formal details):

(i) As a first step we look at individual patient scores, all obtained through binary-choice tasks. We represent each individual data point not as a number (proportion correct), but rather, as a confidence interval on a binomial distribution as follows: we take the couple (n_{XA}, m_{XA}), where n_{XA} is the total number of trials for a patient A on type X sentences (e.g., **+*Movement***, or −***Mood***) and m_{XA} is the corresponding number of successes. We assume independence between trials, therefore interpreting the whole experiment, for a given patient, as repeated Bernoulli trials with probability parameter P_{XA}. We compute a 99% confidence interval for $P_{X,A}$—spanning from Lower to Upper value

---

[5] For simplicity of exposition, we gloss over string-vacuous movement, that is, over movement from VP-internal position into subject position, as well as movement from subject position in relative clauses. These distinctions are suppressed as they are not presently relevant, see Grodzinsky (2004) for further discussion.

$[P_{\mathrm{LXA}}, P_{\mathrm{UXA}}]$.[6] The relevant set of confidence intervals is now computed for the different methods of categorization. Informal, visual inspection of the results provides a preliminary indication as to the categorization of choice.

(ii) To get a numerical evaluation of our contrasts, we apply the appropriate hypothesis testing method at the group level: we assume that for each subject, the binary response (success or failure) forms a set of Bernoulli trials whose probability parameter is a random variable varying from subject to subject following a certain probability law. We then assess the hypothesis that performances on sentence types X, Y, come from the same underlying distribution of individual probability parameters. Since a *t* test approach cannot be used here,[7] we assume that the probability of success follows a probability law given by the parametric family $B(\alpha,\beta)$ (the β-distribution). This form of distribution is capable of a variety of shapes as its parameters vary, so that no severe limitation is imposed on the way the probability fluctuates (Skellam, 1948). Moreover, it has been widely used on biological data (Williams, 1975). Implementation was carried out in *Mathematica*, using the Find Minimum function on the negative log-likelihood to find the parameters of the β-distribution. We validated our implementation by reproducing the numerical results of a published biological test (Williams, 1975).[8]

4.1. Complexity vs. Movement in relative clauses

Applying this two-pronged method, we first compared the categorizations ***Complexity*** and ***Movement*** on the relative clause data set (Fig. 1).

Fig. 1 presents two levels of contrast in English relative clauses. Each sentence contains a main clause (**bolded**), and a relative clause (bracketed, courier subscript). Rows constitute the ***Complexity*** level: the relative can either modify the subject of the main clause (resulting in the more complex Center Embedding **a, b**), or the object (in the less complex Right Branching **c, d**). Columns constitute the ***Movement*** level: in the [bracketed] relative clause, the grammatical function of the head (**the woman, the man**) can either be Subject ◀ (**a, c**) or Object ▶ (**b, d**). In Subject relative constructions (**a, c**), the blue relative head connects to an empty proximal position ◀, located on the left edge of the relative clause. The link <**the woman-◀**> crosses no part of the relative, and is analyzed as $-$***Movement***.[9] In Object-relative constructions (**b, d**), by contrast, the red relative head connects to ▶, a distal, non-left-edge, empty position. The **+*Movement*** link <**the man-▶**>, depicted by the arrow, crosses both the verb and the subject of the relative clause.[10]

The 2×2 data matrix has 64 data points from 23 English speaking Broca's aphasics, with results arranged according to the contrasts given in Fig. 2:

In panel (A), the Center Embedding (perforated red line) vs. Right Branch (full black line) contrast is represented by a vertical display of the confidence intervals for all patients. Each vertical line represents the confidence interval for a given patient. Panel (B) is same as panel (A), this time $-$***Movement*** = perforated red line, +***Movement*** = full black line. Panel (C) represents the β-model for CE (Red) and RB (Black). The *x*-axis represents possible values of the probability parameter, the *y*-axis units are such that the area under the curve is 1. Panel (D) is same as panel (C), this time − ***Movement*** = red, +***Movement*** = black.

The complexity contrast (A and C) discerns no structure; ***Movement*** was vastly superior—the contrast was highly significant (Fig. 2; see Table 1 for all numerical results), indicating that ***Movement***, but not ***Complexity***, is a linguistic factor supported by Broca's region.

5. Mood vs. Movement in actives and passives

A second quantitative comparison that we carried out was between ***Mood*** and ***Movement***, for which we used the active and passive data. Consider what is needed to test this contrast: we must find an array of monoclausal sentences in which word-order changes (i.e.

---

[6] Notice that the confidence interval summarizes information for each patient about both the percentage of successful trials and the overall number of trials. This is necessary in view of the fact that the number of trials for the two branches of a contrast are not always equal in a given experiment (if it was not specifically designed to test the contrast under scrutiny), let alone in different experiments. To display all the results for a given contrast we have chosen to show each patient as a vertical line spanning the corresponding $[P_{\mathrm{LXA}}, P_{\mathrm{UXA}}]$ interval.

[7] We have three reasons to disqualify a *t*-test approach: (A) summarizing the performances of each patient as a percentage of success gives an empirical distribution of numbers in the $[0,1]$ ranges, which does not square with the range of the normal distribution assumption underlying the *t*-test. (B) This move is unwarranted also because the number of trials for different patients (and even for different contrasts for a given patient) is not fixed. This means that the percentage values cannot be pooled safely. (C) More conceptually, we are trying to determine the distribution over the population of Px for a given X, and there is no reason to assume a priori that this distribution is bell shaped (Fig. 4 shows a case in which the best fitting distribution is J-shaped).

[8] See supplementary materials for a formal presentation, implementation, and validation.

[9] The relative pronoun *who* is viewed here as being in between the clauses, belonging to neither; we assume irrelevance of so-called "vacuous movement" for the perspective developed here (or a "±vacuous movement" contrast).

[10] Some data from three additional langugages (Chinese, Japanese and Korean) exist in the data set, but was not included in the analysis for logistical reasons.

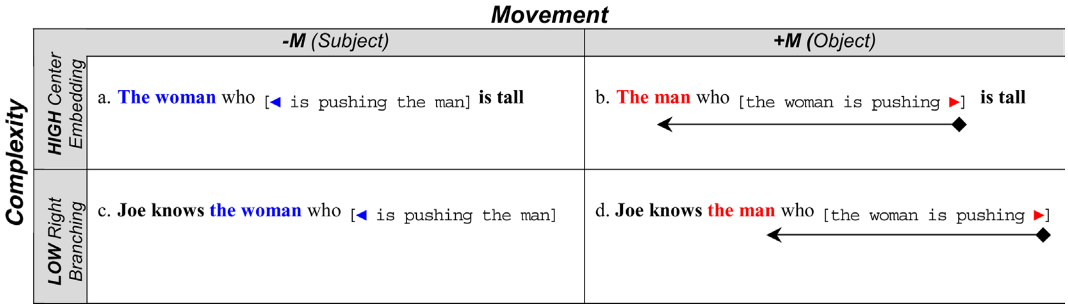

Fig. 1. Movement vs. complexity in relative clauses.

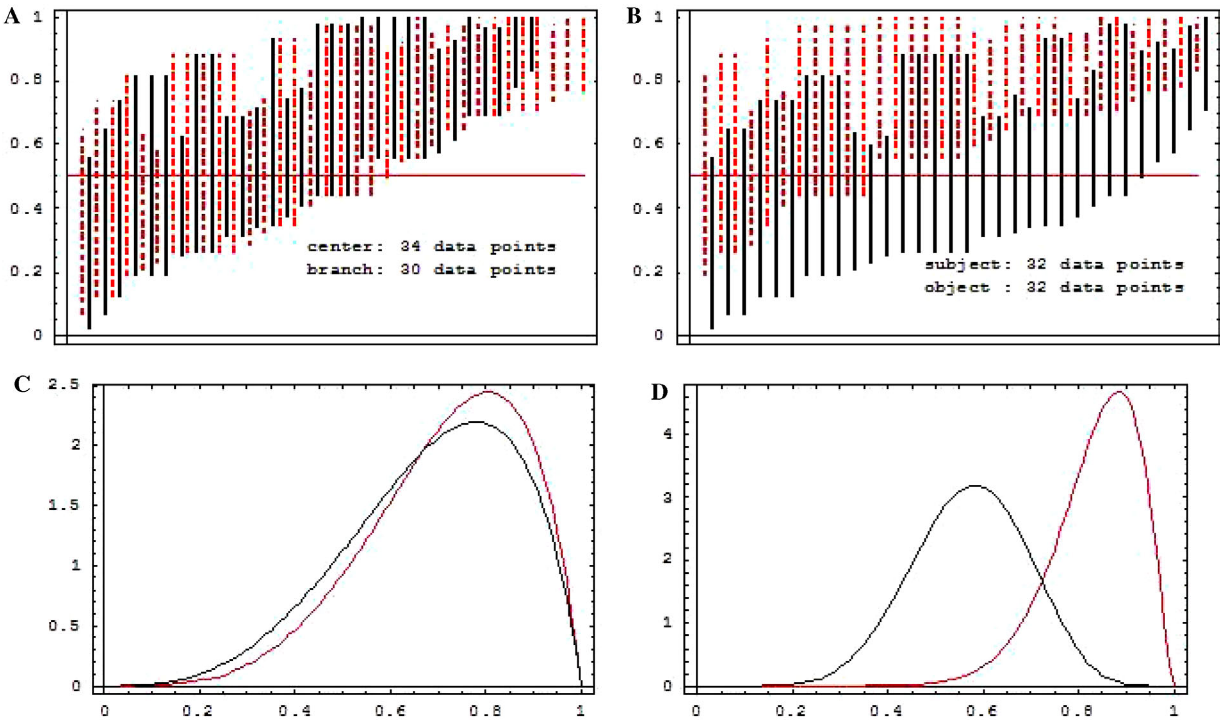

Fig. 2. Individual confidence intervals and β-distribution for relative clauses. (See text for interpretation.)

Table 1
Summary of parameters of the β-models fitted to each contrast, together with ratios and p value of the test for the null hypothesis (i.e., no performance difference between the branches of the contrast)

	Mood vs. movement (English included)				Mood vs. movement (English excluded)				Complexity vs. movement				All movement	
	Passive/Active		+Mov/−Mov		Passive/Active		+Mov/−Mov		HIGH (CE)/LOW(RB)		+Mov/−Mov		+Mov/−Mov	
M	0.713	0.804	0.610	0.849	0.768	0.791	0.594	0.866	0.716	0.687	0.573	0.835	0.594	0.842
$\sum$	0.194	0.164	0.172	0.123	0.168	0.179	0.118	0.110	0.160	0.170	0.119	0.09	0.154	0.117
$\sigma_{\text{easy}}/\sigma_{\text{hard}}$	1.183		1.398		0.938		1.072		0.941		1.322		1.316	
p	0.015		4.7×10^{-13}		0.57		2.9×10^{-15}		0.84		2.4×10^{-8}		$<10^{-16}$	

Movement) and morphological (i.e., ***Mood***) changes are independent of one another. Ideally, this yields a 2×2 design: one axis in which word order is kept constant and morphology is changed; and another axis in which morphology is kept constant and word order is changed.

How can morphological change and word order change be teased apart in mono-clausal sentences? Looking at English (or Hebrew, or Spanish), this task appears nearly impossible: as we pointed out above, English passive constructions contain passive morpholo-

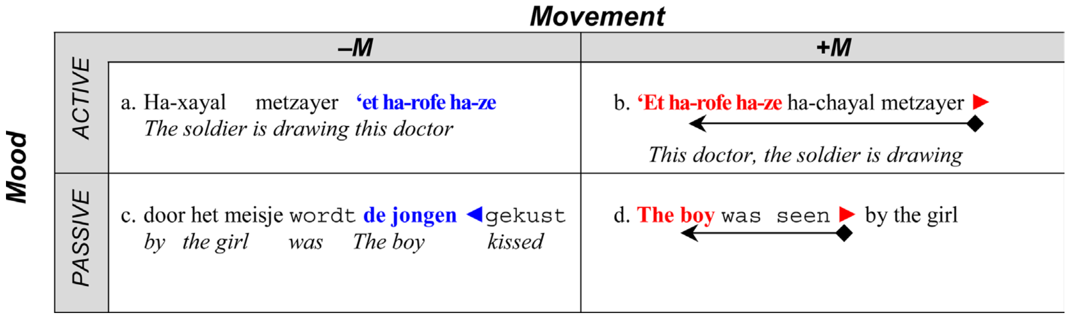

Fig. 3. Movement vs. mood in simple sentences. (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this paper.)

gy, an inserted auxiliary verb, which in most cases are accompanied by a change in word order relative to the active counterpart. English (or Hebrew, or Spanish) passive is thus derived by syntactic movement. These are SVO (Subject–Verb–Object) languages, in which simple active declarative sentences feature the subject on the left of the verb, and the direct object on its right. In passive, morphological changes and auxiliary insertion co-occur with the movement of the object from the right of the verb leftward. The object thereby crosses the verb [__*Saw Bill →Bill was seen*__]. This complex cluster of active/passive differences makes it difficult to tease ***Mood*** apart from ***Movement*** through these constructions in this language (although see Wasow, 1977, for a classical paper on passive constructions that are not derived by movement; see Grodzinsky, Pierce, & Marakovitz, 1991, for an investigation of these in aphasia).

Yet this picture changes if we extend our field of view to other language types. SOV (Subject–Object–Verb) languages like German or Dutch avail us of the desired contrast. There, passivization does not necessarily co-occur with verb crossing by the (internal) argument that becomes the subject of the passive sentence. To illustrate, the Dutch analogs of the English fragments above are [__*heeft het meisje gekust →het meisje wordt*__ *gekust*].[11] *Het meisje* (the girl) maintains its position relative to the participle *gekust*, despite auxiliary change (cf. the bottom left quadrant of Fig. 3). ***Mood*** change in Dutch/German does not necessarily entail word order change (auxiliary notwithstanding). This provides one half of the desired contrast (constant order, varying morphology), which we can now explore quantitatively.

The other half of the contrast (varying order, constant morphology) would be a ***Movement*** contrast that does not entail ***Mood*** change (as reflected in morphological change). This we find in English, Hebrew, Spanish, Korean, and Japanese, where simple active declarative sentences manifest two varieties: either the subject precedes the object (S..O..), or the object precedes the subject (O..S..). This can happen through either scrambling, or topicalization, depending on the language and the operation chosen. Such an order change (in which meaning remains unmodified) is mostly viewed as an instance of ***Movement*** in both SVO and SOV languages (for recent discussion, cf. Fanselow, 2001, for German; Saito & Fukui, 1998, for Japanese). Crucially, while the subject and the object switch positions, the verb remains in active voice, keeping ***Mood*** constant (e.g., Fig. 3 below, top right quadrant). This ***Movement*** operation thus occurs within a single active clause, and without visible morphological or ***Mood*** changes. ***Movement*** in simple sentences, then, can be independent of change in ***Mood***. We used this contrast in our test.

We have thus illustrated how all four possibilities are realized, and it important to note that the syntactic analysis we assume is not controversial. The resulting array is summarized in Fig. 3, with new example sentences. And while there are complications that go beyond the scope of this discussion, our discussion is empirically justified (though somewhat simplified for ease of exposition, see Grodzinsky, in press, for detailed elaboration and justification of some of the necessary linguistic assumptions).[12] Fig. 3 is set up to reflect this distinction: sentences may contain ***Movement*** without passive morphology (cf. top right corner of Fig. 3), as well as Passive morphology without ***Movement*** (bottom left corner).

The same figure presents the logical structure of our second test, which sought to see which distinction is apparent in Broca's aphasia: the traditional active/passive (Mood) distinction or Movement. Fig. 3 thus presents two levels of contrast in monoclausal sentences. Rows constitute the traditional grammar ***Mood*** level:

---

[11] For simplicity, we ignore here an extra complication: SOV configuration in Dutch and German is evident in simple sentence only when there is an auxiliary in addition to main verb (cf. *Bill sah Peter* vs. *Bill hat Peter gesehen*). For our purposes, this issue can be ignored, as the SVO–SOV contrast is maintained throughout our sample. See Grodzinsky, 2004a, for further elaboration.

[12] The position of the subject relative to the *by*-phrase in German can be reversed (i.e., the order *door het meisje wordt de jongen* ◀ *gekust* is also possible). One experiment tested both orders, with the same results (Friederici & Graetz, 1987). The near-perfect performance on both types of sentences, as well as linguistic considerations pertaining to the (adjunct) status of the *by*-phrase, allows us to abstract away from this contrast.

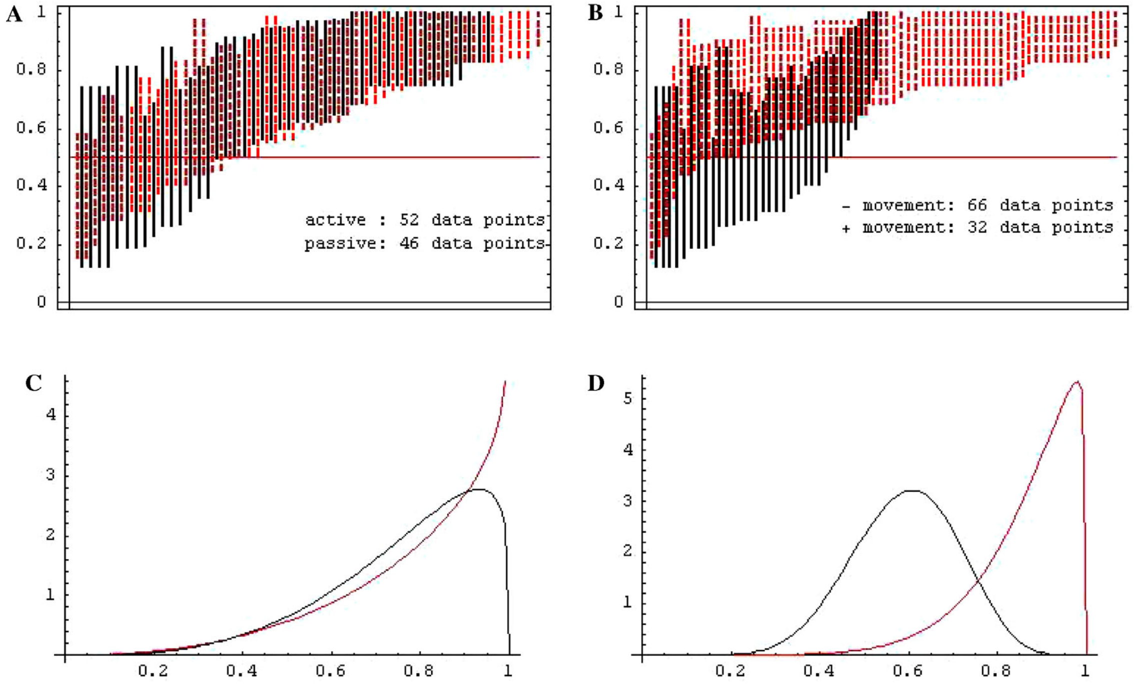

Fig. 4. Individual confidence intervals and β-distribution for non-English active/passive. (For interpretation of the references to colors in this figure legend, the reader is referred to the web version of this paper.)

actives (**a**, **b**) lack passive morphology, and their "logical object" is in Object position. Passives (**c**, **d**) contain an auxiliary verb (e.g., English *is*, Dutch *wordt*), and a passive morpheme (e.g., English-*en*, Dutch *ge-V-t*); the "logical object" is located in Subject position. Columns constitute the ***Movement*** level: −***Movement*** trivial links: (a) Object (blue) is exactly in the position of recipient-of-action in Hebrew active; (**c**) "Logical object" (blue) is adjacent to the position of recipient-of-action in Dutch passive <**de jongen-◀**>. Non-trivial **+*Movement*** links: (b) The arrow-marked link between the topicalized Object (red) and the position of recipient-of-action <ha-rofe, ▶> crosses the Subject and the verb (d) The arrow-marked link <**the boy** – ▶> in the passive crosses the verb.

5.1. The quantitative analysis

The complete active/passive data subset contained 155 points, from 69 Broca's aphasic speakers of five languages (Fig. 3).[13] Recall that in English, ***Movement*** and ***Mood*** mostly overlap.[14] For this reason, we analyzed our data twice—with and without the English data. We first analyzed it as a whole, namely with 57 English data points included. A significance difference was found for both contrasts—namely ***Mood*** and ***Movement*** (Table 1). Next, we excluded the English cases, and only kept data from the four languages that featuring a ±***Movement*** contrast *within* actives, and a ±***Movement*** contrast in passive *within* and *between* languages (98 scores, 42 Broca's speakers of Dutch, German, Hebrew, and Spanish).

Our results are presented in Fig. 4 as follows: Panel (A) the Active (perforated red line) vs. Passive (full black line) contrast is represented by a vertical display of the confidence intervals for all patients. Each vertical line represents the confidence interval for a given patient. Panel (B) is same as panel (A), this time −***Movement*** = perforated red, +***Movement*** = full black. Panel (C) the β-model for Active (Red) and Passive (Black). The x-axis represents possible values of the probability parameter, the y-axis units are such that the area under the curve is 1. Panel (D) is same as panel (C), this time −***Movement*** = red, +***Movement*** = black. The ***Mood*** contrast (A and C) discerns no structure, while the ***Movement*** contrast discerns structure, and yields highly significantly different probability distributions, providing further evidence of the strong relation between a ***Movement*** deficit and Broca's aphasia, and indicating that ***Mood*** is not a natural class vis-à-vis the syntactic comprehension deficit in this syndrome.

To underscore this last point, we present a comparison between English and Dutch/German passive, that is, a subset of the data in Fig. 4 that include **+*Movement*** passives in English ($n = 34$), and the −***Movement*** passives in Dutch/German ($n = 18$). A robust difference is apparent (Fig. 5).

---

[13] Some data from three additional languages (Chinese, Japanese and Korean) exists in the data set, but was not included in the analysis.

[14] Except special cases, such as adjectival passive, cf. Grodzinsky, Pierce and Marakovitz (1991).

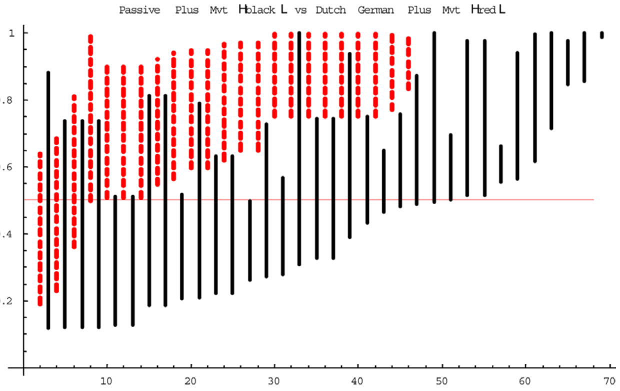

Fig. 5. Individual confidence intervals of ±***Movement*** passive: Dutch/German (perforated red line), vs. English (full black line).

Finally, a combined analysis, in which all our data were pooled together and partitioned by the ±***Movement*** contrast, was highly significant[15] ("all Movement" in Table 1), strengthening the view that the two separately investigated contrasts fall under a single syntactic generalization (e.g., Chomsky, 1995).

6. Remaining variation

While the movement effect is robust, and helps discern structure in the data, variation no doubt exists in our data set, as can be revealed by mere inspection of the individual confidence intervals, or, more formally, by the σ values of each of the β-distributions (Table 1; see Appendix B for the details of the calculation of for this distribution). We do not know whether this variation exceeds the variation documented through other experimental methods, in other populations or cognitive domains. Still, it may be worthwhile to try and investigate the variation in our data in more detail. In a preliminary attempt to do so, we examined the σ values of the different curves generated for the comparisons above. Unstructured variation would lead to similar σ values across curves. If we were to expect structured variation, then harder experimental conditions would likely lead to σ values that are greater than those of the easier conditions, because the latter would be more likely to reach ceiling performance. This means that $\sigma_{\text{hard}}/\sigma_{\text{easy}} > 1$. If our reasoning is valid, then the σ-ratio might be used to point to the correct hard/easy contrast in the data.

The evidence we have in this respect is weak, but suggestive: it appears that +***Movement*** experimental conditions produce data whose variance is greater than the −***Movement*** conditions. That is, in all instances, $\sigma_{+\text{Mov}}/\sigma_{-\text{Mov}} > 1$ (1.398 for the ±***Movement*** contrast in active/passive with English; 1.072 for the same contrast in active/passive without English; 1.322 for the ±***Movement*** contrast in relative clauses; 1.316 for the same contrast for all sentence types). The other contrasts, by comparison, are non-uniform in this respect: the active/passive (with English) case, the "hard/easy" ratio is 1.183; the same comparison without English is <1 (0.938); finally, as is the $\sigma_{\text{CE}}/\sigma_{\text{RB}}$ ratio in the branching contrast of the relative clauses (0.941). These are, admittedly, weak effects, but they do point to a direction consistent to our claims.

7. A concluding remark

The new method for quantitative analysis we propose shows that for the 69 patients we studied, Broca's aphasia leads to a robust ***Movement*** deficit, cutting across other elements that have been thought to generate comprehension difficulties (***Mood***, ***Complexity***). Although the group we have studied is large, it is, of course, difficult to assert that the selection criteria we used are universally valid; yet, they do define a clinical population that exhibits stable behavioral deficits, and that these deficits characterize what clinicians have in mind when they think of Broca's aphasia. Neurologically, then, existent diagnostic methods appear efficacious for this syndrome. Thus, while variability exists—as it does in virtually all areas of biology—the robust structure we uncovered in the data, and its relation to clinical diagnostic tests of Broca's aphasia, are clear.

Still, the variability debate is unlikely to stop here. We would thus like to clarify what we think we did by carrying out this analysis, and what we did not. We begin on the negative side: first, our analysis is entirely empirical, and we have nothing to say regarding conceptual issues in neuropsychology. Second, our analysis has nothing to say about the specificity of deficit in Broca's region (or lack thereof). That is, as we focused on Broca's aphasics, we did not address the question of whether there exist comprehension tests that tease it apart from other syndromes. Hopefully, future work along the same lines will help determine this rather difficult issue. Finally, we do not show that variation does not exist in this syndrome. On the contrary: it is evident in the data, just like in most results in neuroscience. What we did show, however, is how our methods discover highly robust structure in a large data set that was previously said to be resistant to such structural analyses. It is this result and this analysis, and nothing else, that we would now like to bring to the fore.

---

[15] An additional 18 data points, from related syntactic contrasts (Subject- vs. Object-questions, Subject- vs. Object-Topicalization) were included in the last analysis.

Appendix A. The data sources

English active/passive

Badecker, W., Nathan, P., & Caramazza, A. (1991). Varieties of sentence comprehension deficits: A case study. *Cortex, 27*, 311–321.
Berndt, R., Salasoo, A., Mitchum, C. C., & Blumstein, S. E. (1988). The role of intonation cues in aphasic patients' performance of the grammaticality judgement task. *Brain and Language, 34*, 65–97.
Caplan, D., & Futter, C. (1986). Assignment of thematic roles to nouns in sentence comprehension by an agrammatic patient. *Brain and Language, 27*, 117–134.
Grodzinsky, Y., Pierce, A., & Marakovitz, S. (1991). Neuropsychological reasons for a transformational derivation of syntactic passive. *Natural Language and Linguistic Theory, 9*, 431–453.
Hickok, G., & Avrutin, S. (1995). Representation, referentiality, and processing in agrammatic comprehension: Two case studies. *Brain and Language, 50*, 10–26.
Linebarger, M., Schwartz, M., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13*, 361–392.
Ostrin, R. K., & Schwartz, M. F. (1986). Reconstructing from a degraded trace: A study of sentence repetition in agrammatism. *Brain and Language, 28*(2), 328–345.
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*(2), 85–113.
Schwartz, M. F., Saffran, E. M., & Marin, O. S. M. (1980). The word order problem in agrammatism: Comprehension. *Brain and Language, 10*, 249–262.
Shankweiler, D., Crain, S., Gorrell, P., & Tuller, B. (1989). Reception of language in Broca's aphasia. *Language and Cognitive Processes, 4*, 1–33.
Sherman, J. C., & Schweickert, J. (1989). Syntactic and semantic contributions to sentence comprehension in agrammatism. *Brain and Language, 37*, 419–439.

Non-english active/passive

Beretta, A., Harford, C., Patterson, J., & Pinango, M. (1996). The derivation of post-verbal subjects: Evidence from agrammatic aphasia. *Natural Language and Linguistic Theory, 14*, (4), 725–748.
Beretta, A., Schmitt, C., Halliwell, J., Munn, A., Cuetos, F., & Kim, S. (2001). The effects of scrambling on Spanish and Korean agrammatic interpretation: Why linear models fail and structural models survive. *Brain and Language, 79*, 407–425.
Burchert, F., De Bleser, R. & K. Sonntag. (2001). "Does case make the difference?", *Cortex, 37*, 700–703.
Friederici, A. & Graetz, P. A. M. (1987). Processing Passive Sentences in Aphasia: Deficits and Strategies. *Brain and Language, 30*, 93–105.
Friedmann, N., & Shapiro, L. P. (2003). Agrammatic comprehension of simple active sentences with moved constituents: Hebrew OSV and OVS structures. *Journal of Speech Language and Hearing Research, 46*, 288–297.
Hagiwara, H. (1993). The breakdown of Japanese passives and theta role assignment principle by Broca's aphasics. *Brain and Language, 45*, 318–339.
Kolk, H., & van Grunsven, M. (1985). Agrammatism as a Variable Phenomenon. *Cognitive Neuropsychology, 2*(4), 347–384.
Kolk, H., van Grunsven, M. & Keyser, A. (1985). On parallelism in agrammatism. In M. -L. Kean (Ed.), *Agrammatism.* New York: Academic Press.

Relatives

Burchert, F., De Bleser, R. & K. Sonntag. (2001). "Does case make the difference?" *Cortex, 3*, 700–703.
Beretta, A., Harford, C., Patterson, J., & Piñ ango, M. 1996. The derivation of post-verbal subjects: Evidence from agrammatic aphasia. *Natural Language and Linguistic Theory, 14*(4), 725–748.
Beretta, A., Schmitt, C., Halliwell, J., Munn, A., Cuetos, F., & Kim, S. 2001. The effects of scrambling on Spanish and Korean agrammatic interpretation: Why linear models fail and structural models survive. *Brain and Language, 79*, 407–425.
Friedmann, (2000). Moving verbs in agrammatic production. In R. Bastiaanse & Y. Grodzinsky (Eds.) *Grammatical disorders in aphasia: A neurolinguistic perspective.*
Grodzinsky, Y. (1989). Agrammatic comprehension of relative clauses. *Brain and Language, 37*, 480–499.
Hickok, G., & Avrutin, S. (1995). Representation, referentiality, and processing in agrammatic comprehension: Two case studies. *Brain and Language, 50*, 10–26.
Sherman, J. C., & Schweickert, J. (1989). Syntactic and semantic contributions to sentence comprehension in agrammatism. *Brain and Language, 37*, 419–439.

Questions

Hickok, G., & Avrutin, S. (1996). Comprehension of Wh-questions in two Broca's aphasics. *Brain and Language, 52*, 314–327.
Thompson, C. K., Tait, M. E., Ballard, K. J., & Fix, S. C. (1999). Agrammatic aphasic subjects comprehension of subject and object extracted wh-questions. *Brain and Language, 67*, 169–187.

Clefts

Caplan, D., & Futter, C. (1986). Assignment of thematic roles to nouns in sentence comprehension by an agrammatic patient. *Brain and Language, 27*, 117–134.

Appendix B. The methodology

The performance of a subject X on a set of sentences S is described by a couple ($n_{\mathrm{X,S}}$, $m_{\mathrm{X,S}}$) where $n_{\mathrm{X,S}}$ = number of trials and $m_{\mathrm{X,S}}$ = number of correct answers. We assume independence between trials, therefore interpreting the whole experiment, for a given patient, as repeated bernoulli trials with probability parameter $p_{\mathrm{X,S}}$. We then proceed to evaluate $p_{\mathrm{X,S}}$ in light of the observed performances of patient X on the sentences S.

A 99% confidence interval [$p_{\mathrm{LX,S}}$, $p_{\mathrm{UX,S}}$] for $p_{\mathrm{X,S}}$ is obtained by solving the following equations:

$$\sum_{k=0}^{m_{\mathrm{X,S}}} \binom{n_{\mathrm{X,S}}}{k} p\mathrm{UX}^k (1 - p\mathrm{UX})^{n_{\mathrm{X,S}-k}} = .01/2,$$

$$\sum_{k=0}^{m_{\mathrm{X,S}-1}} \binom{n_{\mathrm{X,S}}}{k} p\mathrm{LX}^k (1 - p\mathrm{LX})^{n_{\mathrm{X,S}-k}} = (1 - .01)/2.$$

This approach leads to a first way of assessing the pertinence of a contrast: we simply display confidence intervals for $p_{\mathrm{X,S}}$ for each subject as well as the confidence intervals for $p_{\mathrm{X,\sim S}}$ (where ~S denotes the trial sentences that are not in S). As can be seen in Figs. 2A and B such a simple visual inspection gives a quite striking impression of the relevance (or lack thereof) of a contrast. Thus, a contrast in group behavior between the $p_{\mathrm{X,S}}$ and the $p_{\mathrm{X,\sim S}}$ becomes visually obvious, but a rigorous hypothesis testing framework is preferable.

One could try to summarize each subject's performance by the percentage of success and then analyze the distribution of these numbers, but since the number of trials varies from study to study, serious distortion of the data would likely occur. Indeed, a deviation of 10% from chance in 10 trials does not have the same statistical meaning as a deviation of 10% from chance in 100 trials. The conventional tools for testing the hypothesis that two samples come from the same distribution (e.g., Mann–Whitney or Wilcoxon) are therefore not applicable here and we need a framework which suits the binomial nature of the data, while leaving room for the possibility that the probability coefficient in the corresponding Binomial distribution may vary from subject to subject. Such a framework is provided by the β binomial distribution, a natural generalization of the binomial distribution, in which the probability parameter (fixed in the binomial) is allowed to vary according to a parametric law $B(\alpha, \beta)$. This distribution is capable of a variety of shapes as its parameters vary, so that no severe limitation is thus imposed on the way the probability of success on a given type of sentence fluctuates from subject to subject [see (Skellam 1948)], moreover it has been widely used on biological data [see (Williams, 1975)].

Formally, under the $B(\alpha, \beta)$ model the probability of m successes in n trials is

$$P(\mathrm{Success} = m | \mathrm{Trials} = n) = \binom{n}{m} B(\alpha + m, n + \beta - m)/\mathrm{B}(\alpha, \beta);$$

with

$$B(\alpha, \beta) = \Gamma(\alpha) * \Gamma(\beta)/\Gamma(\alpha + \beta), \Gamma$$

being the usual Gamma function.

In the context of such a model we can, for each set of subject's performance on a set of sentence, determine by maximum likelihood computation the best fitting value for the parameters α, β. Furthermore for the performances of a set of subjects on two set of sentences (e.g., passive/active, or +Movement/−Movement) we can test the null hypothesis that the whole data set comes from one parametrized distribution against the hypothesis that they come from two distinct distribution. Indeed Ignoring constants involving only the observations, the log-likelihood for a set of results$\{(n_1, m_1)\ldots(n_k, m_k)\}$ is given by

$$L(\mu, \theta | \{(n_1, m_1) \ldots (n_k, m_k)\}) = \sum_{j=1}^{k} \left(\sum_{r=0}^{m_j - 1} \mathrm{Log}[\mu + (r * \theta)] + \sum_{r=0}^{n_j - m_j - 1} \mathrm{Log}[1 - \mu + (r * \theta)] - \sum_{r=0}^{m_j - 1} \mathrm{Log}[1 + (r * \theta)] \right).$$

Note that this likelihood depends on the full information concerning the number of trials and of successes of each subject so that no concern over unduly pooling incomparable percentages arises. If $\{(n_1, m_1)\ldots(n_k, m_k)\}$, represents the performances of subjects on one type of sentences and $\{\{(n_{k+1}, m_{k+1})\ldots(n_{k+q}, m_{k+q})\}\}$ their performances on an other type, we can proceed to test the hypothesis that the performances are similar using the fact that if L_0 is the maximal value of

$$L(\mu, \theta | \{(n_1, m_1) \ldots (n_k, m_k), (n_{k+1}, m_{k+q}) \ldots (n_{k+1}, m_{k+1}q)\})$$

and L_1 is the maximal value of

$$L(\mu, \theta | \{(n_1, m_1) \ldots (n_k, m_k)\}) + L(\mu, \theta | \{(n_{k+1}, m_{k+1}) \ldots (n_{k+q}, m_{k+q})\});$$

then an asymptotically valid test of the similarity of performance for the two types of sentences is given by comparing $2(L_1 - L_0)$ with upper percentage points of the c^2 distribution with two degrees of freedom (see Williams, 1975).

See Fig. 2 for the density curves of the probability of success corresponding to various sentence types, and Table 1 for a summary of the significance of the various contrasts we have tested. This procedure was implemented in *Mathematica*, using the FindMinimum function on the negative log-likelihood in order to find the parameters of the β-distribution. We validated our implementation by reproducing the numerical results presented in Williams (1975).

References

Avrutin, S. (2001). Linguistics and agrammatism. *GLOT International, 5*, 3–11.

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D., & Grodzinsky, Y. (2003). The neural reality of syntactic transformations: Evidence from fMRI. *Psychological Science, 14.5*, 433–440.

Berndt, R. S., Mitchum, C. C., & Haedinges, A. N. (1996). Comprehension of reversible sentences in "agrammatism: A meta-analysis. *Cognition, 58*, 289–308.

Caplan, D., & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Science, 22*, 77–126.

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive processes from patterns of impaired performance: The case for single-patient studies. *Brain and Cognition, 5*, 41–66.

Caramazza, A., Capitani, E., Rey, A., & Berndt, R. (2000). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. *Brain and Language, 76*, 158–184.

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

Drai, D. (in press). Evaluating deficit patterns of Broca aphasics in the presence of high inter subject variability. In: Y. Grodzinsky, K. Amunts (Eds.), Broca's Region. New York: Oxford University Press, forthcoming.

Fanselow, G. (2001). Features, θ-roles, and free constituent order. *Linguistic Inquiry, 32*(3), 405–437.

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

Goodglass, H. (1968). Studies in the grammar of aphasics. In S. Rosenberg & J. Koplin (Eds.), *Developments in applied psycholinguistics research*. New York: Macmillan.

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

Grodzinsky, Y. (1984a). The syntactic characterization of agrammatism. *Cognition, 16*, 99–120.

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

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

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

Grodzinsky, Y. (2004). Variation in Broca's region: Preliminary cross-methodological comparisons. Chapter 8 in Lyle Jenkins, ed., *Variation and Universals in Biolinguistics*. Oxford: Elsevier.

Grodzinsky, Y. (2004a). Variation, canonicity and movement. Paper presented at the EURESCO "Science of Aphasia V" Meeting, Potsdam, September.

Grodzinsky, Y. (in press). A blueprint for a brain map of syntax. In Grodzinsky and Amunts, eds.

Grodzinsky, Y., Pierce, A., & Marakovitz, S. (1991). Neuropsychological reasons for a transformational analysis of verbal passive. *Natural Language and Linguistic Theory, 9*, 431–453.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science, 274*, 114–116.

Kolk, H. (2004). Agrammatism. In R. Kent (Ed.), *MIT encyclopedia of communicative disorders*. Cambridge, MA: MIT Press.

Kolk, H. H. J., & Heeschen, C. (1992). Agrammatism, Paragrammatism and the management of language. *Language and Cognitive Processes, 7*, 82–129.

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

Mohr, J. P. (1978). Broca's area and Broca's aphasia. In H. Whitaker (Ed.). *Studies in Neurolinguistics* (Vol. II, pp. 201–235). San Diego: Academic Press.

Röder, B., Stock, O., Neville, H., Bien, S., & Rösler, F. (2001). Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: A functional magnetic resonance imaging study. *NeuroImage, 15*, 1003–1014.

Saito, M., & Fukui, N. (1998). Order in phrase structure and movement. *Linguistic Inquiry, 29*, 439–474.

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

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society Series B (Methodological), 10.2*, 257–261.

Wasow, T. (1977). Transformations and the Lexicon. In P. W. Culicover, Thomas Wasow, & Adrian Akmajian (Eds.), *Formal Syntax* (pp. 327–360). New York: Academic Press.

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

Zurif, E. B. (1995). Brain regions of relevance to syntactic processing. In L. Gleitman & M. Liberman (Eds.) (2nd ed.). *An invitation to Cognitive Science* (Vol. I.). Cambridge, MA: MIT Press.