

NOTES AND DISCUSSION

Broca's Aphasia Is Associated with a Single Pattern of Comprehension Performance: A Reply

Dan Drai

Tel Aviv University, Tel Aviv, Israel

Yosef Grodzinsky

Tel Aviv University, Tel Aviv, Israel, and Boston University School of Medicine

and

Edgar Zurif

Brandeis University and Boston University School of Medicine

Over many years now, we have provided evidence that the cortical area associated with Broca's aphasia sustains operations necessary for the analysis of syntactic constructions that contain displaced constituents. A massive body of empirical data—from comprehension, real-time processing, and grammaticality judgment—shows that Broca's area supports mechanisms involved in the computation of transformational relations (cf. Zurif, 1995; Grodzinsky, 2000). This evidence has been challenged and debated—most recently in the pages of this journal. Thus in 1999 we published the results of a survey of comprehension scores for Broca's patients conventionally selected via the Boston Diagnostic Aphasia Exam (Goodglass & Kaplan, 1972) and other like instruments (Grodzinsky, Piñango, Zurif, & Drai, 1999; GPZD henceforth). We included all the relevant data published between 1980 and 1996 of which we were aware. The picture that emerged from these data was very clear: The Broca's patients as a group performed significantly above chance level in their comprehension of structures without displaced constituents—actives, subject relatives, and subject-clefts; in contrast, they were at chance level in their comprehension of transformationally derived structures (those with displaced phrases), i.e., on passives, object relatives, and object-clefts. For one particular instance of this generalization—the active/passive contrast—there exists a considerable body of experimental evidence, obtained at different times and in different laboratories. It is on this nar-

Authors are listed alphabetically. The preparation of this paper was made possible by NIH Grants DC 00081 and 02984 to the Aphasia Research Center, Boston University School of Medicine, 03660 to Brandeis University, and Israel–U.S. Binational Science Foundation Grant 97-00451 to Tel Aviv University. The authors thank an anonymous reviewer for helpful comments.

Address correspondence and reprints requests to Yosef Grodzinsky at the Aphasia Research Center, 151-A, Bldg. 9, Boston VAMC, 150 S. Huntington Avenue, Boston, MA 02130. E-mail: yoseff@bu.EDU.

rower issue that Berndt and Caramazza (1999) have focused. Ignoring the rest of the evidence, they have claimed that our data are misleading, the result of incorrect patient selection. We countered their complaints and showed that even for the narrow range of data they discuss, they are mistaken. This we have done through a patient-by-patient examination (Zurif & Piñango, 1999) and through a demonstration of an active/passive contrast in the very series of patients they themselves had presented initially (Drai & Grodzinsky, 1999).

Caramazza, Berndt, and their colleagues are back in this issue of *Brain and Language* with new thoughts on the matter. They now propose that the Broca's patients' data point to at least two distinct groups of patients, and they bring forth statistical and empirical arguments for their position. As we show below, however, the underpinnings of their position violate basic principles of parsimony in statistical modeling. Moreover, their empirical support—the putative patients' scores they use—is flawed and arbitrary. We cover each of these points in turn.

STATISTICAL MODELING

When the frequency distribution of some empirical data appears to suggest a unimodal distribution, how do we know that our impression is correct? In our case, how do we decide whether the frequency distribution we have documented for our patients' comprehension performance for passive sentences (GPZD) is unimodal, as we have claimed, or whether it actually masks two or more modes as Caramazza et al. (2001) now propose? True, our patients are all Broca's aphasics, selected on independent grounds (speech production patterns and neuroanatomical and other clinical data). But that does not guarantee that their comprehension profiles will also look alike. What we need is a method that will allow us to determine whether the data all belong to the same pool without invoking external considerations.

This problem of recognizing distinct components within a frequency distribution arises in other areas of science—in domains as different as electrophoresis and exploratory behavior in the rat. In the former case, one must distinguish between different molecular weights given external forces; in the latter, different modes of behavior are sought, aimed at categorizing seemingly uniform behavior. In other words, many areas need a statistical method to discern structure in seemingly uniform data sets. Not surprisingly, a statistical model has been developed to address this problem: *the Gaussian mixture model*.

The problem begins with the observation that a unimodal curve can always be described as consisting of more than one Gaussian. The question is whether a move toward more, rather than fewer, components is *justified*. The fewer components a curve is described to consist of, the more parsimonious the description. So, any n -modal curve can be described as consisting of a mix of $n + 1$ Gaussians. The real question, then, is not about the possibility of such a description, but rather about its justification, that is, whether the additional $(n + 1)$ th Gaussian is not spuriously introduced. To determine this, a statistical method is available. It can tease apart components within a curve and crucially applies regardless of partial overlap between individual Gaussians. This is the model used for the investigation of electrophoresis and of exploratory behavior in rodents, and it is the one that we will use for the analysis of our aphasia data.

Before we present the algorithm for this analysis and report our finding, however, it is important to contrast our proposal with the argument Caramazza and his colleagues try to make. They point out that two curves may be created from any given single curve. This is true, but trivially so: as we have said above, it is always possible to move from n to $n + 1$ Gaussians to describe the frequency distribution of a popula-

tion. The critical question is whether this move is empirically justified. For that, we need a method and a criterion. Caramazza et al. provide none and leave us with the above trivial point, from which no conclusion whatsoever follows. Below, we show how a rigorous statistical analysis is able to determine whether an $n + 1$ components Gaussian mixture model should be preferred to an n components model, given the data; and to forecast our finding, we demonstrate that there is no reason to model the data from passive comprehension in Broca's aphasia with two Gaussians rather than with one. This decisive finding leads to the conclusion that there is no statistical reason to assume two groups of comprehenders.

ANALYSIS: PARAMETER ESTIMATION IN THE GAUSSIAN MIXTURE MODEL

Using an implementation of the Gaussian mixture model (see, e.g., Everitt and Hand, 1981) that has been applied recently to rodent exploratory behavior (Drai et al., 2000a), we examined comprehension error rates in Broca's aphasia to determine whether our patients constitute one group or more. In effect, we tried to determine whether a decomposition of the single Gaussian we observed in our graph (GPZD, 1999, Graph 1) is spurious (Fig. 1).

The parameters in the model are formed by using the expectation–maximization (EM) algorithm. This algorithm estimates the maximum likelihood parameters (proportions, means, and standard deviations) of a mixture with a given number of components (peaks in a graph, which represent different groupings). EM is an iterative algorithm that starts with user-given initial values and incrementally improves the likelihood function until further iterations yield only a negligible improvement. The actual number of components (= groupings in the data) of the model is determined by comparing the maximum likelihood value of an n -components mixture with that of an $(n - 1)$ -components mixture until the increased number of components ("peaks" in the curve) increases the likelihood only marginally. More precisely, the spurious character of the n th component is manifested through the fact that the log

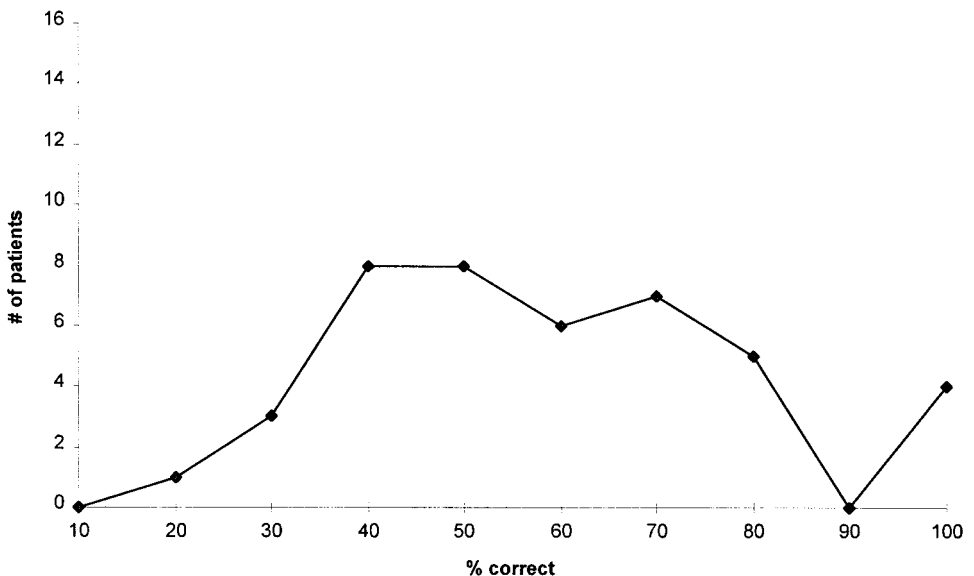


FIG. 1. Passive scores for 42 Broca's aphasics (GPZD, 1999).

of the ratio of the likelihood of n components over that of $n - 1$ components is distributed like a χ^2 with 2 degrees of freedom (see, e.g., Everitt and Hand, 1981). In other words:

If the n th component is spurious then the quantity $I(n) = \text{Log}(p(n)/p(n - 1))$, where $p(k)$ denotes the maximal likelihood under the hypothesis of k components, is distributed like a χ^2 with 2 degrees of freedom (henceforth χ^2_2).

This provides a precise criterion for the determination of the number of components in the mixture. On its basis, we use the following procedure for setting the correct number of components: for each possible value n of the number of components (starting with 1, i.e., a pure Gaussian model), compute with EM the maximal likelihood value $p(n)$ for a model with n components. Then compute $\log(p(n + 1)/p(n))$. Initially, each additional component may improve the likelihood by several orders of magnitude. For a certain number of components k , this improvement ceases to be significant (i.e., $I(k + 1)$ is within the range of χ^2_2) and we can then set k as the correct number of components. This is the empirical criterion that determines whether the addition of a peak is justified. Having found the correct number of modes, k , we then adopt as our model for the data the Gaussian mixture with k components and with parameters (proportion, mean, and SD for each component) that yield maximum likelihood among models with k components.

IMPLEMENTATION

Drai's SEE software (Drai et al., 2000b), which builds on the *Mathematica* programming language, implements this algorithm. We analyzed two sets of data. First, we gave it the passive data from GPZD for which it created the density curve (after removing outliers = 5% extremes; i.e., it computes the curve on the 90% central part of the data).¹

We then asked whether one or two Gaussians fit it best:

1. The best single-peak, one-Gaussian model has a parameter mean of 0.539, an SD of 0.027, and a likelihood of 1.96×10^6 (the product of the values of the density function for each empirical value).

2. The best double-peak, two-Gaussian model has a likelihood of 8.5×10^6 .

3. As our algorithm dictates, we now compute the log of the ratio of the likelihoods. If the number could not reasonably come from χ^2 with 2 degrees of freedom, then we must reject the one-Gaussian model and adopt the two-Gaussian model (or even more, keep on doing the same with three Gaussians, etc.). This ratio is 1.46418.

4. To complete the process, we now compute the χ^2 value ($2df$) of this, which comes to 0.480903. This number is too large to validate rejection of the one Gaussian in favor of the two Gaussian hypothesis. That is, the improvement in likelihood has $p = .48$ of being spurious. We set $k = 1$.

We ran the same algorithm with a "corrected" set of data, namely, those commented by Zurif and Piñango (1999). Nine patients were removed from the GPZD set for reasons given in that paper, to obtain a set that consisted of performances of 33 patients. Here, too, we followed the algorithm as before. The log of the ratio between the likelihoods obtained here was 1.85213, and the χ^2 value ($2df$) of this was $p = .396109$, again, too high to reject the single Gaussian hypothesis. The curve that fits the data best is unimodal, a single Gaussian. Adding a supplementary mode is not warranted by the data.

The conclusion is that from a statistical perspective, the comprehension pattern of

¹ The details of the program can be obtained from the authors upon request.

Broca's aphasics on passive sentences is homogeneous. This method, in fact, is the only statistical analysis that has ever examined this issue empirically.

THE ISSUE OF NUMBER OF TRIALS PER PATIENT

The above analysis leaves two concerns that Caramazza et al. voice: First, they argue that we have ignored the number of trials each patient may have had. This number determines the probability that a particular patient's performance will be around the group mean: The higher the number of trials, the more likely this patient is to be a true representative of the group and, in the case of the passive construction, to score at chance level. This point is conceptually valid, but it is hardly an argument against our analysis. Our analysis can be countered only if it is shown empirically that patients with higher numbers of trials are as likely to be at the tails of the distribution as are patients with fewer trials (or, worse yet, more likely to be at the tails than those with fewer trials). We are unaware of such a demonstration, and in any event, Caramazza et al. (2001) provide none. On the other hand, we show above how analyses of two different patient groups (GPZD's 42 patients, Z&P's corrected group of 33 patients) lead to the conclusion that a unimodal graph fits the data best.

THE ISSUE OF PATIENT SELECTION

The second empirical argument that Caramazza et al. offer is that the patients in our samples were improperly selected. This is untrue. A detailed response to some of their objections has already been provided in Zurif and Piñango (1999). Drai and Grodzinsky (1999) have even demonstrated that in an earlier series of patients reported by Berndt et al. (1996) there are clear and important patterns. Here, we will briefly discuss only Caramazza et al.'s (2001) most current patient samples, specifically, the two groups they refer to as S1 and S2, both of which are different from ours (GPZD, 1999).

To form S1, they drop patients from the GPZD survey that they claim were insufficiently described. As they state, “. . . it was not possible to determine the *N* of trials administered for nine patients included in their samples, either because the original study was unpublished, the patients were not individually identified in the original study, or the number of trials was not included in the original study.” But as can be easily verified, every patient in our original analysis was individually identified in terms of a published study, as were the number of trials he or she received (with one exception, a patient in Badecker et al. (1991), wherein contradictory information was given on the number of trials that patient received). Moreover, they leave out two patients (A and C) from Grodzinsky (1995), but keep two others (B and D)—for no apparent cause. Also Sherman and Schweickert's (1989) study is flatly dismissed, even though all their patients (DM, EM, FC, LD, RD) can be easily identified through other studies in which they participated (e.g., Grodzinsky, Pierce, & Marakovitz, 1991; Hickok & Avrutin, 1995; Grodzinsky, 1995a, 1995b). All this information is available in GPZD's Appendix 1, which Caramazza et al. apparently did not read.

A second reason for excluding patients from S1 has to do with the number of trials each patient received. If there are too few in Caramazza et al.'s estimation, the patient is dropped. On this basis, they reject seven patients from Hagiwara's (1993) study on the grounds that these patients had only 6 trials. But they then include two subjects (AK and GL) from Martin's (1987) study, who were tested with 8 trials each. Why is 8, and not 6, their cutoff point? No reason is given, and so we are left wondering why these last two patients are included in their sample. Also in this vein, there is

an unaccounted discrepancy that these authors do not attempt to reconcile: in our study, as in Berndt et al.'s, no patient had over a hundred trials. In Caramazza et al.'s group three patients do have such large numbers.

As noted above, Caramazza et al. (2001) also discuss another patient grouping, S2, with 33 patients. Here, they include 13 new patients from Benedet, Christiansen, and Goodglass's (1998) study. Had Caramazza et al. delved into the details of this paper seriously, they would have found that it used the materials from Goodglass, Christiansen, and Gallagher (1993). This would have forced them to either include this latter study, which they did not, or reject both, as they should have. The active/passive test of both studies is a giveaway: Quite simply, it does not include reversible pictures. One look at the stimulus example given (Goodglass et al., 1993, p. 406, Fig. 2) shows that for the passive sentence "the boy is found by the dog" two pictures are provided, one showing a dog finding a man and the other showing a dog finding a child. Given this contrast, no syntactic analysis is necessary for a correct answer. Indeed, most patients performed well above chance in this condition, having a major effect on the character of Caramazza et al.'s 52 grouping. Caramazza et al. also include 4 patients studied by Druks and Marshall (1991). But as we have pointed out elsewhere, these investigators used ungrammatical word strings to test comprehension, and to make matters worse, 2 of their 4 patients were not native speakers of Hebrew, the language in which they were tested (Grodzinsky, 1992; Zurif, 1996). In short, Caramazza et al. (2001) do not provide a credible database for their analysis. No amount of statistical analysis can transcend irrelevant data, nor can an unjustified dismissal of good data be condoned.

CONCLUSIONS

We have countered Caramazza et al.'s arguments and are now in a position to conclude. Caramazza et al. have challenged a small part of the database regarding Broca's area and Broca's aphasia. For aphasia, this database consists of comprehension findings for a wide variety of constructions in several languages (cf. Grodzinsky, 1995a, 1995b, 2000), of real-time processing results (cf. Zurif, 1995), and of grammaticality judgment scores (Grodzinsky & Finkel, 1998). All these data suggest that the receptive impairment in Broca's aphasia has to do with the analysis of displaced constituents and that Broca's area is the locus of processes that carry out this task. In light of all this evidence, Caramazza et al.'s (2001) insistence on individual differences in Broca's aphasia wrongly diverts attention from the important regularities in the data. Contrary to their claims, variability in processing rates (e.g., Swinney et al., 1996) and chance factors do not obscure the solid generalization concerning the neuroanatomy of syntactic comprehension that has emerged from the study of Broca's aphasia.

There is yet another disadvantage to Caramazza et al.'s focus on variation. In forming their arguments, they examine only the clinical syndrome of Broca's aphasia, namely, the functional impairment, and not its lesion localizing value. The critical fact is, however, that the syndrome of Broca's aphasia is routinely associated—virtually always—with left anterior inferior cortical damage. So the particular syntactic comprehension deficit described for Broca's aphasia is not simply a fact about the isolability of a linguistic constituent. It is also a fact about the neuroanatomical underpinnings of the syntactic operation affected. It ties the construction of syntactic dependencies to left inferior frontal cortex. And as it happens, converging evidence has emerged from work in other research areas as well. A special syntactic involvement for the area associated with Broca's aphasia has lately been shown by imaging analyses of

normal processing (e.g., Caplan et al., 1998); one fMRI study, in particular, ties this anterior area to the construction of syntactic dependencies, that is, to just those operations identified in Grodzinsky's descriptive generalization (Cooke et al., 1999). Even electrophysiological data have started to converge on the importance of this region for syntactic processing.

All this valuable information is ignored by Caramazza et al., whose paper contains neither arguments nor facts that would lead to a change in view. The debate is concluded, pending new and credible empirical evidence.

REFERENCES

- Badecker, W., Nathan, P., & Caramazza, A. (1991). Varieties of sentence comprehension deficits: A case study. *Cortex*, **27**, 311–321.
- Benedet, M. J., Christiansen, J. A., & Goodglass, H. (1998). A cross-linguistic study of grammatical morphology in Spanish- and English-speaking agrammatic patients. *Cortex*, **34**, 309–336.
- Berndt, R. S., & Caramazza, A. (1999). How “regular” is sentence comprehension in Broca's aphasia? It depends on how you select the patients. *Brain and Language*, **67**, 242–247.
- 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., Alpert, N., and Waters, G. (1998) Effects of syntactic structure and propositional number on patterns of regional blood flow. *Journal of Cognitive Neuroscience*, **10**, 541–552.
- Caramazza, A., Capitani, E., Rey, A., & Berndt, R. S. (2001). Agrammatic Broca's aphasia is not associated with a single pattern of comprehension performance. *Brain and Language*, **76**, 158–184.
- Caramazza, A., Berndt, R. S., Basili, A. G., & Koller, J. J. (1981). Syntactic processing deficits in aphasia. *Cortex*, **17**, 333–348.
- Cooke, A., Zurif, E., DeVita, C., McSorley, C., Grossman, M., & others (1999). Functional neuroimaging of sentence comprehension. *Cognitive Neuroscience Society Abstracts*, 1999:51.
- Drai, D., & Grodzinsky, Y. (1999). Comprehension regularity in Broca's aphasia? There's more of it than you ever imagined. *Brain & Language*, **70**, 139–143.
- Drai D., Benjamini, Y., & Golani, I. (2000a). Statistical discrimination of natural modes of motion in rat exploratory behavior. *Journal of Neuroscience Methods*, **96**, 119–131.
- Drai, D., Golani, I., & Benjamini, Y. (2000b). *SEE: A tool for the visualization and analysis of rodent exploratory behavior*. Tel Aviv University.
- Druks, J., & Marshall, J. C. (1991). Agrammatism: An analysis and critique, with new evidence from four Hebrew-speaking aphasic patients. *Cognitive Neuropsychology*, **8**, 415–433.
- Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.
- Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Philadelphia: Lea & Febiger.
- Goodglass, H., Christiansen, J. A., & Gallagher, R. (1993). Comparison of morphology and syntax in free narrative and structured tests: Fluent vs. nonfluent aphasics. *Cortex*, **29**, 377–407.
- Grodzinsky, Y. (1992). Writing should come after reading: Reply to Ouhalla. *Linguistische Berichte*, **139**, 197–201.
- Grodzinsky, Y. (1995a). A restrictive theory of agrammatic comprehension. *Brain & Language*, **51**, 26–51.
- Grodzinsky, Y. (1995). Trace deletion, theta roles, and cognitive strategies. *Brain and Language*, **51**, 469–497.
- Grodzinsky, Y. (2000). The neurology of syntax. *Behavioral & Brain Sciences*, **23**, 1–71.
- Grodzinsky, Y., & Finkel, L. (1998). The neurology of empty categories: Aphasics' failure to detect ungrammaticality. *Journal of Cognitive Neuroscience*, **10**, 281–292.
- Grodzinsky, Y., Pierce, A., & Marakovitz, S. (1991). Neurophysiological reasons for a transformational derivation of syntactic passive. *Natural Language & Linguistic Theory*, **9**, 431–453.
- Grodzinsky, Y., Pinango, M., Zurif, E., & Drai, D. (1999). The critical role of group studies in neuropsychology: Comprehension regularities in Broca's aphasia. *Brain & Language*, **67**, 134–147.
- Hagiwara, H. (1993). Comprehension of passive in Japanese aphasics. *Brain & Language*, **45**(1), 247–263.

- Hickok, G., & Avrutin, S. (1995). Comprehension of Wh-questions in two Broca's aphasics. *Brain and Language*, **50**, 10–26.
- Martin, R. C. (1987). Articulatory and phonological deficits in short-term memory and their relation to syntactic processing. *Brain and Language*, **32**, 159–192.
- Sherman, J. C., & Schweickert, J. (1989). Syntactic and semantic contributions to sentence comprehension in agrammatism. *Brain and Language*, **37**, 419–439.
- Swinney, D., Zurif, E. B., Prather, P., & Love, T. (1996). Neurological distribution of processing resources underlying language comprehension. *Journal of Cognitive Neuroscience*, **8**, 174–184.
- Zurif, E. (1995). Brain regions of relevance to syntactic processing. In L. Gleitman & M. Liberman, Eds., *Invitation to cognitive sciences, Vol. I*, 2nd ed. Cambridge, MA: MIT Press.
- Zurif, E. (1996). Grammatical theory and the study of sentence comprehension in aphasia: Comments on Druks and Marshall (1995). *Cognition*, **58**, 271–279.
- Zurif, E., & Piñango, M. (1999). The existence of comprehension patterns in Broca's aphasia. *Brain and Language*, **70**, 133–138.