COMPREHENSION AND PRODUCTION IN EARLY LANGUAGE DEVELOPMENT: COMMENTS ON SAVAGE-RUMBAUGH ET AL.

Elizabeth Bates

University of California, San Diego

La Jolla, California 92093

Monographs of the Society for Research in Child Development

Serial No. 233, Vol. 58, Nos. 3–4, 1993, pp. 222–242

COMPREHENSION AND PRODUCTION IN EARLY LANGUAGE DEVELOPMENT: COMMENTS ON SAVAGE-RUMBAUGH ET AL.

Elizabeth Bates

Savage-Rumbaugh and her colleagues have provided us with yet another ground-breaking investigation into the linguistic abilities (or "quasi-linguistic abilities" — see below) of our nearest phylogenetic neighbor, the chimpanzee. Their monograph begins with some brief but useful reviews of the primate language literature, and the literature on early comprehension and production of language in human children. The authors document the peculiar bias toward production and the relative neglect of comprehension that have characterized the child language literature, and they ask a perfectly reasonable question: If we want to understand what an organism *knows* about language, isn't comprehension a better place to start? And if we want to compare knowledge of language in two related species, how can we draw any firm conclusions if our work is based exclusively on what the animal can produce?

With this foundation, Savage-Rumbaugh et al. go on to present (in exquisite methodological and empirical detail) a longitudinal study comparing the development of word and sentence comprehension in a human child (Alia) and a bonobo chimpanzee (Kanzi), raised and tested in settings that are as comparable as ethics and common sense will allow. In contrast with many previous studies of primate language, blind testing procedures are used to ensure against the kind of cueing that proved to be responsible for the supposed linguistic and arithmetic comprehension of the infamous horse Clever Hans. In all honesty, I cannot think of anything else that the authors could have done to convince their audience that this is a fair test of the hypothesis that apes are capable of at least some language comprehension, at both the lexical and the structural level.

I, for one, am convinced. Indeed, it seems fair to conclude from this work that the bonobo (or at least one bonobo) is capable of language comprehension that approximates (in level if not detail) the abilities of a human 2-year-old on the threshold of full-blown sentence processing. I will therefore devote my comments to the implications of this important and startling result, with particular emphasis on the relationship between comprehension and production in human children.

Humans I: Why is there so little research on comprehension?

I agree wholeheartedly with Savage-Rumbaugh et al. that our field has neglected the early stages of language comprehension in human children. With very few exceptions (most of them listed in their review), most of what we know about the first stages of language development is based upon the child's stumbling efforts to produce and reproduce meaningful speech. The authors put their fingers on the main reason why comprehension receives so little attention: it is notoriously difficult to study in this age range. Behavioral methods all require the child to pay attention, follow instructions and carry out some kind of task set up by the adult — whether it is picture pointing, choosing an object from an array, looking back and forth at slides, or carrying out a series of commands. Children under the age of two are often (very often) unwilling to cooperate in a study of this kind. Hence the proportion of false negatives is unacceptably high.

As a direct result of this compliance issue, the internal reliability of comprehension tests tends to be unacceptably low. In several previous studies (e.g. Bates, Benigni, Bretherton, Camaioni & Volterra, 1979; Bates, Bretherton and Snyder, 1988), we have examined the correlations *between* alternative tests of comprehension at 10, 13, 20 and 28 months of age (e.g. a multiple-choice test using real objects; a picture-pointing task; a task in which children are asked to follow simple commands). We have also looked at test-retest and split-half correlations *within* several of our laboratory measures. Results so far have been very discouraging. By 28 months of age, it is possible to administer structured tests like the Peabody Picture Vocabulary Test to normally developing children (Dunn and Dunn, 1981). We can also obtain systematic and reliable data from 2-year-old children using experimental measures of sentence comprehension (e.g. Bates, MacWhinney, Caselli, Devescovi, Natale & Venza, 1984). Before that age, comprehension tasks are highly unreliable. At 13 months, correlations among laboratory measures of word comprehension hover in the +.10 - +.50 range (compared with correlations between +.50 - +.75 for laboratory measures of word production). By 20 months of age, internal correlations for production have gone up still further, but reliability coefficients for laboratory measures of comprehension are still very low (in the non-significant +.18 - +.28 range for many tasks).[1] Spearman's Law states that no measure can correlate with another measure at a level higher than it correlates with itself. Hence, if we try to examine the cognitive, social or neurological predictors of language comprehension in this age range, the unreliability of our comprehension measures sets an absolute ceiling on the power and reliability of any results that we might hope to obtain. This reliability problem extends beyond studies using the correlational method. Developmental researchers sometimes forget that traditional experimental designs (e.g. age by experimental condition) are also limited by the internal reliability of the dependent variable. If we use a behavioral measure of comprehension as outcome variable in a study of this kind (e.g. a study comparing novel-word learning under different conditions), it will be difficult to obtain reliable between-group results with a dependent variable that barely correlates with itself.

In short, if we compare comprehension (based on what we tell the child to do) with production (based on what the child chooses to do, of her own free will), we run the risk of underestimating the former and overestimating the latter. To be sure, some progress has been made in the assessment of early language comprehension. Three methods come to mind: (1) improved uses of parental report to tap into "language comprehension in the wild"; (2) new preferential-looking paradigms that minimize behavioral demands on the child; (3) event-related brain potentials recorded while children are listening to linguistic stimuli. Each of these techniques has some

[1] Internal reliabilities tend to be substantially higher for parental report measures of com-prehension — but as we shall see shortly, parental report has limitations of its own.

real advantages over traditional methods of laboratory testing, but each has some serious disadvantages that make them unsuitable for a study of the kind that Savage-Rumbaugh and her colleagues have presented here. Let us consider each of these innovations in turn.

Parental report. We have known for some time that parental diaries are the best way to measure emerging language abilities in the first two years of life (Dromi, 1987; Darwin, 1971; Leopold, 1949; Stern, 1965). After all, parents are with the child in many different situations, including all those highly predictable routine settings that are the birthplace of early words (e.g. feeding, bathing, going to bed — see Chapter 3 in this monograph). Typically, a child who is capable of producing 20 - 40 words will show no more than 5 of these words to an itinerant researcher visiting the home for two hours with a camera crew. When the same child is brought into an unfamiliar laboratory setting, our estimates of language production may be even lower. For example, we combined parental report with laboratory and home observations in our longitudinal study of 27 infants, from 10 to 28 months of age (Bates et al., 1988). According to detailed interviews with the parents (a predecessor to our current parental report scale — see below), these children had an average expressive vocabulary of 12 words at 13 months of age, with a range from 0 - 45. By contrast, we only observed an average of only 1.69 distinct words in a two-hour combination of laboratory and home testing at the same age level, with a range from 0 - 9. In the same interview, parents reported an average receptive vocabulary of 48 words, with a range from 17 - 97. By contrast, performance for the group as a whole was barely above chance in a 3-way multiple-choice test for the comprehension of familiar object names. To be sure, these different sources of information were significantly correlated, in relatively specific patterns (i.e. parental reports of comprehension with comprehension testing; parental reports of production with observations of production). But it should be clear that parental report offers a much broader view of early words. As we pointed out in the 1988 monograph, parental report yields information about what the child knows how to do, while observations give us a robust estimate of what the same child is willing to do in a short period of time.

Of course, one might argue that parental reports reflect wishful thinking, compared with foolproof laboratory evidence. That is, we can trust positive findings in the laboratory (e.g. the child really did say 'tiger') more than we can trust parental report (e.g. Mom thinks that the child can say 'tiger'). However, we were surprised to find that our 13-month parental report measures were much better long-term predictors of language performance in the laboratory! For example, parental reports of comprehension at 13 months were significantly related to performance on the Peabody Picture Vocabulary Test at 28 months of age (+.56, $p < .01$); by contrast, our 13-month laboratory measure of word comprehension bore no significant relation to any of our later comprehension tests. In fact, this finding follows in a straightforward fashion from Spearman's Law, i.e. from the fact that parental reports of early language have higher internal reliability than corresponding laboratory measures.

As a result of this study (and others that yield a similar result), we joined efforts with a large group of developmental researchers in the U.S. and Europe, trying to find a valid and reliable way to bottle the diary study for mass production. This effort has resulted in a product called the ***MacArthur Communicative Development Inventories*** (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick & Reilly, in press). The CDI contains two overlapping instruments: the Words and Gesture Scale (for normal infants from 8 - 16 months of age, or their developmental equivalent in retarded populations) and the Words and Grammar Scale (for toddlers between 16 - 30 months, or their developmental equivalent in retarded populations). The Words and Gesture Scale taps into word comprehension and production through a 396-word checklist (derived by trial and error through several successive studies, with parents adding new words on every round). There is also a 67-item checklist for different aspects of communicative and symbolic gesture. The Words and Sentences Scale includes a 680-item checklist for the evaluation of word production, followed by a series of checklists that measure aspects of early grammar. The CDI has now been normed with a sample of more than 1,800 healthy children between 8 - 30 months of age, and numerous studies are now available demonstrating the reliability and validity of the various subscales. For example, Dale and his colleagues have shown that the grammatical complexity subscales correlate with laboratory measures of grammar (including MLU and a standard index of syntactic complexity), with coefficients ranging from +.60 to +.86 depending on the age of the sample and the outcome measure in question (Dale, 1991; Dale & Bates, in preparation; Dale, Bates, Reznick & Morisset, 1989). In short, these measures work very well for the global assessment of lexical, gestural and grammatical ability before 30 months of age. The success of these measures reflects three rules that we had to learn the hard way: (1) only ask about current behaviors (retrospective reports of language development have proven unreliable); (2) only ask about newly emerging behaviors (i.e. aspects of language and communication that are still so new that parents can keep track); (3) rely on recognition memory instead of recall, avoiding any need for parents to make complicated inferences that they are not trained to carry out (hence the reliance on checklists instead of "fill in the blank").

At this point, we are convinced that parental report is the best way to obtain a global estimate of language comprehension for children in the first and second year of life. In particular, parental report permits us to circumvent the problem of internal reliability in behavioral measures of comprehension. At the same time, the CDI has serious limitations, all of which are relevant to the Savage-Rumbaugh et al. study. First, the parents of normally developing children can only keep track of word comprehension up to about 16 months of age. After that point, they throw up their hands and say "I don't know, s/he seems to understand just about everything." Second, parental report can only tell us about comprehension in context, where the child has many additional sources of information available to support comprehension of words (e.g. parental gestures and tone of voice; familiar objects and events). We have no way of knowing how well the same child would perform out of context, in a blind testing situation. Third, because it seems to be all but impossible for parents to keep track of comprehension after 16 months, we cannot use this method to assess the emergence of receptive grammar. And there is, of course, a fourth limitation that is particularly important for researchers interested in the symbolic abilities of non-human primates: Many readers are still very skeptical of parental report, and any benefit of the doubt they might be willing to lend in a study of human beings would be denied to researchers working with another species.

Preferential looking. As Savage-Rumbaugh et al. point out in their review of the child comprehension literature, Golinkoff and Hirsh-Pasek have had considerable success in the use of a preferential-looking measure to assess early comprehension of grammar (Golinkoff, Hirsh-Pasek, Cauley & Gordon, 1987;

Hirsh-Pasek & Golinkoff, 1991). Like the research team who developed the CDI to "bottle" parental report, these two researchers have spent many years perfecting the preferential-looking paradigm, holding all the usual disadvantages of that measure to an absolute minimum. Problems of wandering attention have been minimized by monitoring the child's eye movements carefully, and presenting stimuli only when it is clear that the child is attending to the display. To insure attention to the two screens and the linguistic stimuli that emit from a central speaker, they have developed a kind of "sound and light show" that precedes each language trial. Children are briefly familiarized with both the visual display and the relevant sounds before the crucial trials, i.e. the trials in which measures are taken of preferential looking to the picture that "matches" the input sentence (e.g. "Big Bird is hugging Cookie Monster!" presented simultaneously with two scenes, one of Big Bird hugging the Monster, another with the Monster hugging Big Bird). Various steps are also taken to eliminate any possibility of unconscious cueing by the parent (e.g. parents wear earphones that play music to mask verbal input to the child, and they are unable to see the screen even though the child is sitting on the parent's lap with a perfect view). Thanks to all these hard-won precautions, Golinkoff and Hirsh-Pasek and their colleagues have been able to demonstrate that several aspects of phrase and sentence comprehension precede production of the same forms, by several weeks or months. In my view, this work has decisively settled an old controversy (noted by Savage-Rumbaugh in their review) on the relative timing of comprehension and production at the sentence level.

Although these improved preferential-looking methods have provided important new evidence concerning the average onset time for receptive grammar in healthy human children, they too have serious limitations. First, the preferential-looking method rests on a critical assumption: that the child will look longer at a visual stimulus that "matches" the auditory input. The fact that Golinkoff and Hirsh-Pasek have obtained good results so far suggests that this assumption is valid — at least for a significant majority of the children in each study. However, it is worth pointing out that there is a large literature on preferential looking in children under six months of age that makes exactly the opposite assumption: Children will look longer at novel or surprising stimuli that do not match their expectations (Spelke, 1992). If both tendencies are present in individual children, we have to worry about the meaning of null results (i.e. those stimuli or those age groups that do not produce preferential looking at the "match").

Second, although this method works well for group studies, it has proven impossible (at least so far) to adapt the preferential-looking technique for use with individual children (Golinkoff, personal communication). In the experiments that they have conducted to date, Golinkoff and Hirsh-Pasek can obtain no more than 4 - 6 crucial target trials for any linguistic contrast. Although the results are quite reliable at the group level, the predicted pattern (i.e. preferential looking at the pictures that match the language input) is typically displayed by only 2/3 of the children, with looking biases that average around 66% for individual subjects. It should be clear why this kind of hit rate would be unacceptable for individual case studies. To reach significance by a binomial test in a 2-choice situation with six trials, an individual child must perform perfectly on 6/6 trials. And yet the base-rate performance observed in these studies averages four trials out of six — despite all the authors' heroic efforts. If the number of trials could be extended through multiple sessions with the same child, this limit could be overcome. However, this would only provide us with information about a few linguistic contrasts, leaving us with little information about the rest of language comprehension.

These two problems probably suffice to explain why Savage-Rumbaugh and company have avoided preferential looking in favor of traditional behavioral measures (where the probability of getting the right answer by chance is considerably smaller). But there is another reason as well: preferential looking works best with docile children who are willing to sit on the parent's lap for at least 15 minutes, looking at pictures. From the authors' description (and my own observations as a visitor in their laboratory many years ago), chimpanzees are considerably less cooperative than many human children with attention deficit disorders. Barring heavy use of sedatives, I doubt that the preferential-looking measure would prove reliable for the healthy, active, mobile chimpanzee — which brings me to the next point.

Event-related brain potentials (ERP). In the last few years, two laboratories have begun to apply electrophysiological techniques to the study of early language comprehension (Molfese, 1990; Mills, Coffey & Neville, in press-a, in press-b). Electric potentials are measured at the scalp, from infants and children who are wearing a special "hat". Behavioral demands on the child are relatively minimal: They need to cooperate by keeping the hat in place, and they must attend to a series of auditory stimuli played over speakers (holding relatively still while each stimulus is played). Nothing else is required. These techniques have at least four advantages over the preferential-looking technique: (1) although both measures require the child to pay attention to linguistic stimuli, electrophysiological measures make no assumptions about direction of preference (i.e. the assumption that children prefer to look at a "match"); (2) the preferential-looking method usually requires a complex coordination of visual and auditory stimuli, in contrast with ERP studies where auditory stimuli alone are sufficient; (3) because the presentation of stimuli is so straightforward in ERP studies, most individual children can handle a relatively large number of trials; (4) whereas preferential looking only provides a single, relatively unstable dependent variable (percent time looking at the "matching" display), ERP studies elicit a complex, multi-dimensional dependent variable, with variations in timing, amplitude, polarity and scalp distribution. Because of the complex and multidimensional nature of the ERP, it is possible (at least in principle) for electrophysiological researchers to detect fine-grained discriminations among linguistic stimuli and/or the characteristics of individual subjects.

One of the most important ERP studies of early language comprehension is a recent paper by Mills et al. (in press-b), who examined the ERPs associated with familiar vs. unfamiliar words in infants between 13 and 20 months of age. Among other things, these authors have discovered a particular component of the ERP that distinguishes between the two word types. This "comprehension wave" is present in children whose parents report relatively high levels of comprehension; it is absent in children whose parents relatively little evidence for word comprehension. The topological distribution of this component changes with development, from a bilateral distribution that is larger over posterior regions of the brain, to a distribution that is more prominent over left frontal cortex. Most important for our purposes here, there are differences in the shape and distribution of the "comprehension wave" in children who are also able to produce those words. In other words, the ERP can be used to distinguish between "comprehenders" and "producers" during the first two years of life, suggesting that different brain systems are involved in these two aspects of early language — a difference that can be detected in a "pure" comprehension task, with no overt motor response.

The Mills et al. paper raises an obvious question for Savage-Rumbaugh et al.: Given the similarities that have been observed in the comprehension abilities of child and chimpanzee, would we also expect similar patterns of brain activity in response to known and unknown words? On the other hand, in view of the fact that Alia goes on to achieve much higher levels of word and sentence production that Kanzi has displayed so far, would we find tell-tale differences in the brain waves associated with known and unknown words, differences that predict their later differentiation in language output? I would love to know that answer to this question — but I suspect it will be long in coming. It has taken Molfese, Neville, Mills and their colleagues many years to develop the normative information required for the interpretation of brain waves in human infants and children. When this technique is applied to another species, with a very different brain, all this norming and validation would have to begin from scratch. At the end of this process, it would be difficult to say whether we are seeing the "same" brain waves in response to the "same" linguistic stimuli. I also suspect that the compliance problems associated with preferential-looking studies would also plague electrophysiological research with chimpanzees. Mills et al. have used every trick in the book to obtain cooperation from their human subjects — and yet, it is still the case that many children absolutely refuse to wear the hat, or will not sit still long enough to permit collection of passive ERPs without motor artifacts. I sometimes wonder whether we are working toward a rich theory of language comprehension in the docile child — a theory that may not extend to their more rambunctious peers. Obviously the same problem is multiplied a hundredfold when the technique is applied to chimpanzees.

I have reviewed these three techniques for two reasons: (1) to amplify the review of human language comprehension presented by the authors of this monograph, and (2) to prepare the way for some of the additional findings that I will review below, on the dissociation between comprehension and production in the first two years of life. I have not reviewed these techniques to suggest that Savage-Rumbaugh et al. could have done a better job. Indeed, I am persuaded that the behavioral techniques applied in this study are exactly right for the questions that they ask. They have compensated for the notorious problems of reliability and compliance by collecting a very large sample, under blind testing conditions, with careful training to maximize attention and minimize extraneous sources of misunderstanding. Detailed information is provided on the conditions that surround every success, and every failure. The only imperfections that I can detect in this study are those imposed by an imperfect reality.

Humans II: Why do comprehension and production come apart?

In this section, I want to review four new findings on the relationship between comprehension and production in human children, and then go on to explore the implications of these findings for the comparison of apes and children.

Dissociations between comprehension and production in normal children. For many years, parents and psycholinguists have known about children who appear to understand far more than they are able or willing to say (Goldin-Meadow, Seligman & Gelman, 1976; Benedict, 1979; Snyder, Bates & Bretherton, 1981; Bates et al., 1988). The prevalence of this pattern in normal children has now been established on a large scale in the MacArthur CDI norming study (Fenson et al., in press; see also Thal, Fenson and Bates, in preparation). Figure C1 illustrates the relationship between receptive and expressive vocabularies in a cross-sectional sample of 659 children whose parents filled out the Word and Gestures Scale of the CDI when their children were between 8 and 16 months of age. The overall correlation between comprehension and production in this sample was +.65 (+.45 when age is partialled out). This is actually somewhat higher than the coefficients reported in some of our other studies (e.g. Snyder et al., 1981). Nevertheless, we can see from Figure C1 that a significant number of children are producing very little meaningful speech despite receptive vocabularies of 150 words or more. As Bates et al. (1988) have shown, this pattern of dissociation also tends to be a stable characteristic of individual children between 13 and 28 months of age, starting at the single-word level and continuing into the early stages of grammar. The reasons for this robust dissociation are still unclear, but a possible explanation has begun to emerge from studies of abnormal populations.

Dissociations between comprehension and production in late talkers. Specific language impairment is, by definition, a syndrome in which children fall at least one standard deviation below the mean on a criterial language measure, in the absence of any frank evidence for mental retardation or neurological disorders that might account for the delay (Miller, 1991; Tallal, 1988). Most clinicians are unwilling to make a diagnosis of SLI before the child is 3 - 4 years of age. The term "late talkers" is reserved for children who fall far below the mean on measures of expressive language before the 3-year point. For example, Thal and her colleagues define late talkers as children with expressive vocabulary scores in the bottom tenth percentile on the CDI Words and Sentences Scale between 18 - 28 months of age. Although this is an intentionally "pre-clinical" label, several recent studies have shown that approximately 40% of the late-talker population goes on to qualify for a diagnosis of SLI (Rescorla & Schwartz, 1990; Thal, 1991; Thal, Tobias & Morrison, 1991; Whitehurst, Fischell, Arnold & Lonigan, 1992).

Late talkers are defined by their delays in language production. As Thal and her colleagues have shown, there is enormous variability in receptive language skills within the late-talker group, ranging from children who are equally delayed in comprehension and production, to children who are indistinguishable from their chronologically matched controls on measures of comprehension despite severe delays in language output. In other words, a subset of the late-talker population presents with an extreme variant of the comprehension/production dissociation described above. A follow-up study of ten late talkers by Thal, Tobias and Morrison (1991) suggests that comprehension is an excellent predictor of recovery from expressive language delays. That is, children who are building their receptive knowledge of language on a normal schedule have a much better chance of catching up with their age mates across the board in the next 6 - 12 months. *What the child knows is ultimately a better predictor of language ability than what s/he actually does between 18 and 28 months of age.* Although we cannot generalize instantly from language-delayed children to the chimpanzee, I believe this finding is relevant to the work that Savage-Rumbaugh et al. have presented here. It underscores the importance of cross-species comparisons based on levels of language comprehension, to supplement the usual comparisons based on language production. This is particularly true in the developmental range that Savage-Rumbaugh et al. have explored in the present study.

Cognitive correlates of comprehension vs. production. In the last twenty years, a host of studies have appeared examining the non-linguistic correlates of early language development (for reviews, see Bates and Thal 1991; Bates, Thal and Janowsky, 1992; Bates, Thal & Marchman, 1991). Many of these studies were initially inspired by Piaget's theory of the passage from sensorimotor to symbolic cognition in the first three years of life (e.g. Piaget, 1962), and researchers began with the expectation that language milestones would follow across-the-board changes in many different cognitive domains. That is not at all how things turned out. Instead, most researchers this field have arrived at a consensus: Specific linguistic skills are associated with specific abilities outside of language, in a many-to-many relationship that is quite different from the one-to-one stage shifts predicted by orthodox Piagetian theory. Bates et al. (1979) refer to this as the "local homology model"; Fischer and his colleagues describe similar results under the term "skill theory" (Fischer, 1980); Gopnik and Meltzoff refer to the same conclusions as "the specificity hypothesis". Though details differ, the basic underlying intuition is the same in most modern studies of language and cognitive development during the infant and preschool years. For example, we now know that the onset of word comprehension between 8 - 10 months of age is correlated with a host of changes inside and outside of language proper: imitation of novel models, gestural routines (e.g. "bye-bye") and other communicative gestures (e.g. giving, showing and pointing), changes in the ability to recognize a category shift in a passive categorization task, a decline in the ability to recognize phonetic distinctions that are not in the child's native language, advances in causal analysis and the ability to use tools. Around 12 - 13 months of age, the onset of naming in the vocal modality (e.g. pointing and saying "Doggie!") is accompanied by a much more specific set of changes outside of language, in particular the use of conventional gestures to recognize or "name" familiar objects (e.g. putting a telephone receiver to the ear; touching a shoe to the foot; touching a comb briefly to the top of the head). Between 16 - 20 months, two dramatic changes take place within expressive language: a rapid acceleration in rate of language development, and the onset of multiword speech. These two changes are correlated with several developments outside the boundaries of language, including reorganizations in symbolic play (in particular, a shift from "one gesture" to "two gesture" sequences in doll play), changes in active categorization tasks (e.g. successive touching of all the objects in one category, followed by successive touching of objects in a different category), and a shift in the kind of planning that a child displays in block construction. Some researchers have also reported a correlation between the "grammar burst" that usually takes place between 20 - 30 months of age, and a marked increase in the use of conventionally ordered scripts in doll play (e.g. giving teddy bear a bath, with each action occurring in the right order).

All of these correlational studies involve children who are developing on a normal schedule. What happens when components of language fall out of synchrony? In particular, who gets custody of the cognitive correlates when comprehension and production come apart? So far, our studies of normal children and late talkers yield one very clear conclusion: *In almost every case, the child's level of performance in non-verbal cognitive tasks is best predicted by his/her current level of language comprehension.* In other words, cognitive measures are tied most closely to what the child knows about language. Expressive deficits may reflect impairments or delays in some more peripheral aspect of language processing — which brings us to the next point.

Comprehension and production in children with early focal brain injury. We have known for some time that children can recover from brain injuries that would result in irreversible damage in an adult (e.g. Hecaen, 1976; Woods and Teuber, 1978). This does not mean, however, that the brain is totally plastic and equipotential for language or any other cognitive function (Thal, Marchman, Stiles, Aram, Trauner, Nass & Bates, 1991; Satz, Strauss & Whitaker, 1990; Bates et al., 1992; Stiles and Thal, in press). Children with early focal brain injury do display initial problems with language, spatial cognition, affect and attention — that is, with those behavioral domains that are mediated by specific brain regions in the adult. This suggests that there are indeed some initial biases in the human brain. Under normal circumstances, these initial biases lead to the familiar patterns of brain organization that are described in the neuropsychological literature for adults. However, when these "default" conditions do not apply, the infant brain can find alternative neural and/or behavioral solutions, resulting in unusual forms of brain organization that are not usually seen in normal or brain-damaged adults. This seems to be particularly true for language; indeed, most children with early unilateral brain lesions go on to achieve levels of language performance that are indistinguishable from normal on almost every measure (Vargha-Khadem, Isaacs, Van Der Werf, Robb & Wilson, 1991; Vargha-Khadem, Isaacs, Papaleloudi, Polkey & Wilson, 1991; Aram, Holland, Locke, Plante & Tomblin, 1992).

This evidence for the plasticity of language in the human brain is of considerable interest in its own right. However, I want to focus here on those initial biases that have been found in the first years of language learning — with special reference to the brain regions associated with deficits in language comprehension and production. In mature right-handed adults, lesions to anterior (pre-Rolandic) and posterior (retro-Rolandic) areas of the left cerebral cortex tend to result in qualitatively different forms of language breakdown: (1) anterior lesions are associated with non-fluent speech with relatively preserved comprehension at the clinical level (i.e. Broca's aphasia); (2) posterior lesions are associated with fluent but empty speech, marked by word-finding deficits, substitution errors, and mild to severe impairments in comprehension (i.e. Wernicke's aphasia). If the initial delays observed in infants with focal brain injury follow the adult pattern, then we should expect the following patterns in early language development: (1) more severe deficits in expressive language following left anterior injury (i.e. the Broca hypothesis), and (2) more severe deficits in receptive language following left posterior injury (i.e. the Wernicke hypothesis).

In fact, both of these hypotheses have been overturned in our recent prospective studies of language development in children with unilateral brain lesions acquired prelinguistically, i.e. before six months of age (Marchman, Miller & Bates, 1991; Thal et al., 1991). First, results suggest that <u>all</u> children with early focal brain injury are at risk for language delay in the early stages, regardless of size, side or site of lesion. Second, children with lesions extending into left posterior cortex display more severe and persistent delays in expressive (but not receptive) language across this period — directly contradicting both the Broca and the Wernicke hypothesis. Third, receptive deficits were actually more common in children with right-hemisphere damage — direct evidence against the Wernicke hypothesis. These patterns do not occur in our studies of older children with the same etiology; indeed, most of our older children are performing within the normal range. We tentatively suggest that the most intense period of recovery from language delay takes place between 1 - 5 years of age. Furthermore, the regions that mediate language acquisition in the first years of life are not necessarily the same regions that mediate processing and maintenance of language in the adult.

These results for language contrast markedly with our studies of spatial cognitive development in the same population (Stiles-Davis, 1988; Stiles & Nass, 1991; Stiles and Thal, in press). In fact, spatial cognitive deficits (though subtle and less persistent) do bear a systematic relationship to the brain-behavior correlations observed in adults. In line with recent studies of visual analysis in brain-damaged adults, LH injuries in 3 - 12-year-old children result in an *analytic deficit*, i.e. problems with the extraction of "local" perceptual details within a complex visual pattern. RH injuries in the same age range result in an *integrative deficit* (where details are intact but the global configuration is impaired). This suggests that plasticity for language may be greater than plasticity observed with phylogenetically older cognitive systems.

We have proposed two working hypotheses to unify some of our findings for early language and spatial cognition. They are worth reproducing here, because they may be relevant to the comprehension/production disparities observed in other species.

(1) Comprehension as sensory integration. In research on language breakdown in adults, RH lesions do have some effect on complex aspects of discourse processing, e.g. the ability to tell a coherent story, or understand the point of a joke (Brownell, Potter, Bihrle & Gardner, 1986; Gardner, Brownell, Wapner & Michelow, 1983). However, RH lesions typically do not lead to deficits in the comprehension of individual words. And yet, our infant work suggests that RH lesions are sometimes associated with delays in word comprehension. Why should this be the case? To understand this paradox, we have to remember that one-year-old infants are learning to comprehend words for the first time. For adults, comprehension of familiar words is an automatic process, one that takes place without awareness and with very little effort. By contrast, one-year-old infants are still in the process of "cracking the code". For these children, word comprehension may be viewed as a form of multi-modal problem solving, requiring the integration of many different sources of information including gesture, facial expressions, tone of voice, and a host of situational cues (e.g. we are having breakfast now). The adult literature on spatial cognitive deficits suggests that the right hemisphere plays a particularly important role in the integration of sensory information. Stiles' research suggests that the same right-hemisphere bias is operating in early childhood. We propose that sensory integration also plays a particularly important role in the first stages of language comprehension, when children have to use many different sources of information to figure out what words mean. If it is the case that right-hemisphere tissue plays a privileged role (though not an exclusive role) in sensory integration, then we might expect a correlation between right-hemisphere damage and delays in the initial stages of language comprehension.

(2) The local detail hypothesis. Based on a large body of neuropsychological research with adults and children, we propose that the sensory regions of left posterior cortex are particularly important for the extraction of sensory detail. Furthermore, this regional specialization for "local detail" holds for both visual and auditory stimuli. Now, why should a deficit in sensory detail affect production more than comprehension? At first glance, this appears to be a contradiction, because we have always assumed that comprehension is a product of sensory processing, while production relies more on motor factors. For adults who have already acquired their language, this may be true. However, we should remember that children between 0 - 2 years of age are learning to produce their language for the first time. Among other things, this means that they must extract enough perceptual detail from the linguistic signal to support construction of motor templates for production. Our point is really a very simple one: *Perception-for-production requires more sensory detail than perception-for-understanding.*

Let me offer one simple example to illustrate this point. Imagine a 14-month-old infant sitting in her stroller at the zoo. Daddy rolls the child in front of a cage that contains huge long-necked creatures munching away at the lower boughs of a tree. Daddy points to these surprising creatures and says "Cindy, look at the giraffes!" How much acoustic detail does the child need to remember or to learn the word "giraffe" in this situation? In fact, she may get by with nothing more than a salient piece of the word (e.g. something like "uh-RA"). However, the situation changes when Cindy goes home to Mom and tries to tell about her day, reproducing "uh-RA" in a vain effort to talk about that animal with a long neck. A reproduction like "uh-Raff" might be sufficient to do the job (particularly in view of the fact that Mommy knows her child has just returned from the zoo). But "uh-RA" is not sufficient. Back to the drawing board ...

Of course, we could equate the sensory demands on comprehension and production by presenting words in isolation, with no context of any kind. Under these conditions, any differences that we observed between comprehension and production could be blamed on motor demands (the usual suspect in cases of expressive language delay). However, this kind of disembodied speech is rare in the first stages of language learning, and there is no reason to believe that it plays a serious role in the learning process. We agree with Savage-Rumbaugh and her colleagues on the role of familiar routines and contextual support in early language learning. Under these conditions (i.e. "language in the wild"), comprehension and production differ markedly in their reliance on perceptual vs. contextual information. This means, in turn, that there may be proportional differences in the contribution of brain regions that specialize in perceptual analysis (i.e. extraction of sensory detail) and contextual analysis (i.e. integration of information across and within modalities).

In our work with human children, we have proposed that different forms of language delay may result from differential patterns of brain maturation, and/or from subtle deficits at some level of neural computation (i.e. fast vs. slow; detail-oriented vs. integrative units). Notice that we have not invoked <u>any</u> domain-specific language organs to account for these differential patterns of language breakdown. At the same time, we disavow old-fashioned "general cognition" accounts for language delay. There is no such thing as "vanilla cognition". The human brain is a complex and highly plastic computational organ. There are variations in computational style and computational power from one region to another, from one layer to another within a single region, and from cell to cell. Small quantitative variations in computing power (local or distributed) can have important effects on the nature of learning and the way that problems are ultimately solved. And yet none of this requires us to presuppose a rigid, content-specific and highly specialized blueprint for language, or any other aspect of cognition (for a more detailed version of this argument, see Churchland and Sejnowski, 1992). Instead, there may be very indirect routes to "default" brain organization in children and adults, and many alternative routes are possible when default conditions do not hold.

What are the implications of this view for children vs. chimpanzees? Savage-Rumbaugh and her colleagues have shown that chimpanzees are capable of *some* symbolic ability (though not as much as Alia shows by the end of the study), and *some* grammatical ability (though not enough to support full-fledged parsing of a complex sentence). Furthermore, the chimpanzee's level of language ability appears to be better in comprehension than it is in production. Within comprehension, there are subtle differences between child and chimpanzee in item difficulty. For example, Alia performs better with phrasal compounds (e.g. "Get the ball and the banana"), whereas Kanzi has an edge on items with recursion. In my view, this complex pattern of quantitative and qualitative variation between species cannot be explained by postulating a language organ that is present in Alia but absent in Kanzi.

I think that it is time for us to abandon the idea that our brain is organized around content-specific faculties (for language, music, faces, etc.). Is the chimpanzee brain different from the human brain? Of course it is. Without question, these differences are responsible for the presence of full-fledged language in humans, and the absence of anything but "quasi-language" and "weak symbolic capacity" and "rudiments of grammar" in the chimpanzee. However, I believe these differences are due to the computational properties of neural systems that *indirectly* support language learning and language use in human beings. As we have seen, the areas that support language learning are not the same as the areas that support maintenance and use of fluent language in a mature adult. Furthermore, studies of brain activity during language processing in the adult suggest that many different regions are active when language is in use, with different patterns of activation depending on the task (e.g. word comprehension, covert word production, categorization of words in a novel task, categorization of words in a familiar task, translation from one language to another, judgments of grammaticality, studies of priming between word pairs — for reviews see Damasio, 1989; Kutas & Kluender, 1991; Petersen, Fiez & Corbetta, 1992). In other words, even within the human species we find little evidence for a circumscribed language organ. The whole brain participates in language, although the pattern of participation that we see varies depending on the task at hand, and some regions are clearly more important than others.

In a species with a quantitatively and qualitative different brain, we should not be at all surprised to find quantitative and qualitative variations in language ability. Savage-Rumbaugh et al. have provided evidence for both. They are cautious, even scrupulous in their efforts to specify where the species differ and where they overlap. In my view, this is a model for comparative research in cognitive neuroscience. The 21st century is upon us. It is time to abandon phrenology and faculty psychology in favor of dynamic, quantitative models of brain and behavior. Savage-Rumbaugh and her colleagues have encountered opposition and skepticism for many years, because their readers (being human beings) are loath to abandon the idea that we are "special", separate, qualitatively different and unquestionably better than the humble chimpanzee. Of course their opponents are right in one respect: humans are certainly better at language! But "better" is a relative term. The Berlin Wall is down, and so is the wall that separates man from chimpanzee. We are going to have to learn to live with relative differences and permeable borders. It will be hard, but I believe the world will be the better for it.

REFERENCES

Aram, D., Holland, A., Locke, J.L., Plante, E., & Tomblin, J.B. (1992, November). *The biological basis of developmental language disorders*. Miniseminar presented at the Annual Convention of the American Speech-Language-Hearing Association, San Antonio.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy.* New York: Academic Press.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms.* New York: Cambridge University Press.

Bates, E., MacWhinney, B. Caselli, C.M., Devescovi, A., Natale, F., & Venza, V. (1984). A cross-linguistic study of the development of sentence interpretation strategies. *Child Development, 55,* 341-354.

Bates, E. & Thal, D. (1991). Associations and dissociations in child language development. In J. Miller (Ed.), *Research on child language disorders: A decade of progress.* Austin, TX: Pro-Ed.

Bates, E., Thal, D., & Janowsky, J. (1992). Early language development and its neural correlates. In I. Rapin and S. Segalowitz (Eds.), *Handbook of neuropsychology, Vol. 7: Child neuropsychology.* Amsterdam, The Netherlands: Elsevier.

Bates, E., Thal, D., & Marchman, V. (1991). Symbols and syntax: A Darwinian approach to language development. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch and M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development.* Hillsdale, NJ: Erlbaum.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6,* 183-200.

Brownell, H.H., Potter, H.H., Bihrle, A.M., & Gardner, H. (1986). Inference deficits in right-brain-damaged patients. *Brain and Language, 27,* 310-312.

Churchland, P. & Sejnowski, T. (1992). *The net effect.* Cambridge, MA: MIT Press/Bradford Books.

Dale, P. S. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech and Hearing Sciences 34,* 565-571.

Dale, P. & Bates, E. (in preparation). *The relation between lexical and grammatical development from 16 to 30 months.*

Dale, P., Bates, E., Reznick, S. & Morisset, C. (1989). The validity of a parent report instrument of child language at 20 months. *Journal of Child Language 16,* 239-249.

Damasio, A. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33,* 25-62.

Darwin, C. (1971). A biographical sketch of an infant. (Reprinted from *Mind 2.285,* 292-294, 1877). In A. Bar-Adon (Ed.), *Child Language: A book of readings.* Englewood Cliffs, NJ: Prentice-Hall.

Dromi, E. (1987). *Early lexical development.* New York: Cambridge University Press.

Dunn, L., & Dunn, R. (1981). *Peabody Picture Vocabulary Test — Revised.*

Fenson, L., Dale, P., Reznick, J.S., Thal, D., Bates, E., Hartung, J., Pethick, S. & Reilly, J. (in press). *The MacArthur*

*Communicative Development Inventories: User's guide and technical manual***.** San Diego: Singular Publishing Group.

Fischer, K. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87*(6), 477-531.

Gardner, H., Brownell, H.H., Wapner, W., & Michelow, D. (1983). Missing the point. The role of the right hemisphere in the processing of complex linguistic materials. In E. Perceman (Ed.), *Cognitive processing in the right hemisphere.* New York: Academic Press.

Goldin-Meadow, S., Seligman, M.E.P., & Gelman, R. (1976). Language in the two-year-old. *Cognition, 4,* 189-202.

Golingkoff, R.M., Hirsh-Pasek, K., Cauley, K.M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14,* 23-45.

Hecaen, H. (1976). Acquired aphasia in children and the ontogenesis of hemispheric functional specialization. *Brain and Language, 3,* 114-134.

Hirsh-Pasek, K. & Golinkoff, R.M. (1991). Language comprehension: A new look at some old themes. In N. Krasnegor, D. Rumbaugh, R. Schiefelbusch and M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development.* Hillsdale, NJ: Erlbaum.

Kutas, M. & Kluender, R. (1991). What is who violating? A reconsideration of linguistic violations in light of event-related potentials. *Center for Research in Language Newsletter Vol. 6, No. 1.* University of California, San Diego.

Leopold, W.F. (1949). *Speech development of a bilingual child: A linguist's record.* Evanston, IL: Northwestern University Press.

Marchman V., Miller, R., & Bates, E. (1991). Babble and first words in children with focal brain injury. *Applied Psycholinguistics, 12,* 1-22.

Miller, J. (1991). Research on language disorders in children: a progress report. In J. Miller (Ed), Research on child language disorders: A decade of progress (pp. 3-22). Austin, TX: PRO-ED.

Mills, D., Coffey, S. and Neville, H. (in press-a). Language acquisition and cerebral specialization in 20-month-old children. *Journal of Cognitive Neuroscience.*

Mills, D., Coffey, S. and Neville, H. (in press-b). Changes in cerebral organization in infancy during primary language acquisition. In G. Dawson and K Fischer (Eds.), *Human behavior and the developing brain.* New York: Guilford Publications.

Molfese, D. (1990). Auditory evoked responses recorded from 16-month-old human infants to words they did and did not know. *Brain and Language, 38,* 345 - 363.

Petersen, S.E., Fiez, J.A., & Corbetta, M. (1991). Neuroimaging. *Current Opinion in Neurobiology, 2, 217-222.*

Piaget, J. (1962). *Play, dreams and imitation.* New York: Norton.

Rescorla, L. & Schwartz, E. (1990). Outcome of toddlers with specific expressive language delay. *Applied Psycholinguistics, 11,* 393-408.

Satz, P., Strauss, E. & Whitaker, H. (1990). The ontogeny of hemispheric specialization: Some old hypotheses revisited. *Brain and Language, 38,* 596-614.

Spelke, E.S., Breinlinger, K., Macomber, J. & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99, 605-632.*

Snyder, L., Bates, E., & Bretherton, I. (1981). Content and context in early lexical development. *Journal of Child Language, 8,* 565-582.

Stern, C. (1965). *Die Kindersprache. Eine psychologische und sprachtheoretische Untersuchung [von] Clara und William Stern.* (Unveränderter reprographischer Nachdruck der 4., neubearb. Aufl., Leipzig 1928). (Child language. A psychological and linguistic study. [Unrevised reprographic reproduction of new 4th edition, Leipzig 1928].). Darmstadt: Wissenschaftliche Buchgesellschaft.

Stiles-Davis, J. (1988). Spatial dysfunctions in young children with right cerebral hemisphere injury. In J. Stiles-Davis, M. Kritchevsky & U. Bellugi (Eds.), *Spatial cognition: Brain bases and development.* Hillsdale, NJ: Erlbaum.

Stiles, J. & Nass, R. (1991). Spatial grouping activity in young children with congenital right- or left-hemisphere brain injury. *Brain and Cognition, 15,* 201-222.

Stiles, J. & Thal, D. (in press). Linguistic and spatial cognitive development following early focal brain injury: Patterns of deficit and recovery. In M. Johnson (Ed.), *Brain development and cognition: A reader.* Oxford: Blackwell Publishers.

Tallal P. (1988). Developmental language disorders. In J. Kavanagh & T. Truss (Eds.), Learning disabilities: Proceedings of the National Conference (pp. 181-272). Parkton, MA: York Press.

Thal, D. (1991). Language and cognition in late-talking toddlers. *Topics in Language Disorders, 11,* 33-42.

Thal, D., Fenson, L., & Bates E. (1992). *Variations in rate of language development.* Manuscript in preparation.

Thal, D., Marchman, V., Stiles, J., Aram, D., Trauner, D., Nass, R., & Bates, E. (1991). Early lexical development in children with focal brain injury. *Brain and Language, 40,* 491-527.

Thal, D., Tobias, S., & Morrison, D. (1991). Language and gesture in late talkers: A one-year follow-up. *Journal of Speech and Hearing Research, 34:3,* 604-612.

Vargha-Khadem, F., Isaacs, E., Van Der Werf, S., Robb, S., Wilson, J. (1991). Development of intelligence and memory in children with hemiplegic cerebral palsy. *Brain, 115,* 315-329.

Vargha-Khadem, F, Isaacs, E.B., Papaleloudi, H., Polkey, C.E., & Wilson, J. (1991). Development of language in 6 hemispherectomized patients. *Brain, 114,* 473-495.

Whitehurst, G., Fischell, J., Arnold, D. & Lonigan, C. (1992). Evaluating outcomes with children with expressive language delay. In S. Warren & J. Reichle (Eds.), *Causes and effects in communication and language intervention* (pp. 277-324). Baltimore: Paul Brookes.

Woods, B. & Teuber, H. (1978). Changing patterns of childhood aphasia. *Annals of Neurology, 3,* 273-280.

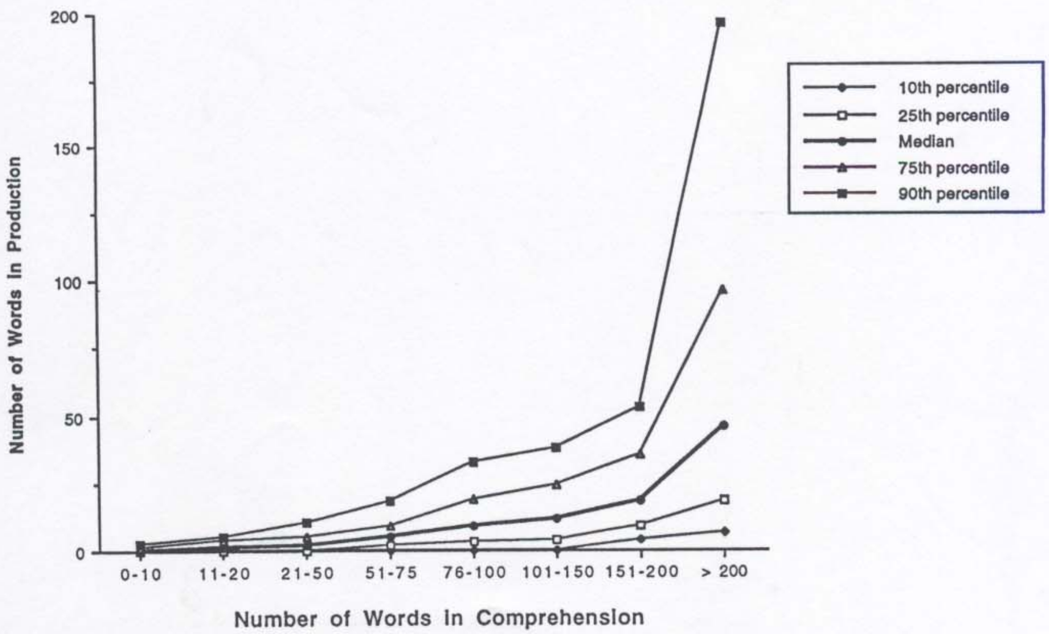

FIG. C1.—Production by comprehension level