

Mental representation of symbols as revealed by vocabulary errors in two bonobos (*Pan paniscus*)

Heidi Lyn

Received: 29 April 2006 / Revised: 17 March 2007 / Accepted: 19 March 2007 / Published online: 14 April 2007
© Springer-Verlag 2007

Abstract Error analysis has been used in humans to detect implicit representations and categories in language use. The present study utilizes the same technique to report on mental representations and categories in symbol use from two bonobos (*Pan paniscus*). These bonobos have been shown in published reports to comprehend English at the level of a two-and-a-half year old child and to use a keyboard with over 200 visuographic symbols (lexigrams). In this study, vocabulary test errors from over 10 years of data revealed auditory, visual, and spatio-temporal generalizations (errors were more likely items that looked like sounded like, or were frequently associated with the sample item in space or in time), as well as hierarchical and conceptual categorizations. These error data, like those of humans, are a result of spontaneous responding rather than specific training and do not solely depend upon the sample mode (e.g. auditory similarity errors are not universally more frequent with an English sample, nor were visual similarity errors universally more frequent with a photograph sample). However, unlike humans, these bonobos do not make errors based on syntactical confusions (e.g. confusing semantically unrelated nouns), suggesting that they may not

separate syntactical and semantic information. These data suggest that apes spontaneously create a complex, hierarchical, web of representations when exposed to a symbol system.

Keywords Categorization · Error analysis · Bonobos · Mental representation · Evolution of language

Introduction

For many years, errors in language use have been employed to clarify features of mental architecture (e.g. Anwar 1979; Crowson 1994; Cutler 1982; Fromkin 1971, 1973, 1980, 1988; Gelman et al. 1989; MacKay 1980). When “mis-speaks” (slips of the tongue, the pen, or in the case of sign language or keyboard users, slips of the hand) occur, such errors are presumed to reveal the existence of an underlying domain of conceptual similarity in the mind of the speaker (Berg 1996; 2005; MacKay 1980; Nelson 1974, 1977). Fromkin (1971, 1973, 1980), in many reviews of speech-error research, detailed errors in every level of human language (e.g. phonological, segmental, morphological, semantic and grammatical) and utilized these errors to argue for the existence of these levels in the mind of the speaker.

Errors have also been utilized to clarify categorical representation in mind of the language user (Gelman et al. 1989). However, categorization is a large field and has therefore been studied by an array of techniques, including sorting and attentional paradigms (e.g. Levy and Schlesinger 1988; Madole and Oakes 1999; Markman 1989; Rosch 1988) and word association tasks. These word association tasks reveal that humans not only make categorical responses but also responses with spatio-temporal associations. For example, dark-night, chair-sit, milk-carton (e.g. Nelson 1977).

Electronic supplementary material The online version of this article (doi:10.1007/s10071-007-0086-3) contains supplementary material, which is available to authorized users.

H. Lyn
Language Research Center,
Georgia State University, 3401 Panthersville Rd,
Atlanta, GA 30034, USA

H. Lyn (✉)
Sea Mammal Research Unit, Gatty Marine Lab,
School of Biology, University of St. Andrews,
St. Andrews, Fife KY16 8LB, UK
e-mail: heidilynphd@aol.com

Additionally, errors have been used to clarify syntactical representations. Children's overregularization errors (for example, "goed" for "went") were the means by which researchers discovered that children's acquisition of grammar was based on generalizations or "rules" rather than on learning narrow associations (e.g. Chomsky 1969; Tomasello 2003).

When studying other linguistic modes such as sign language or writing, errors have revealed evidence that there may be different underlying representational structures depending on such linguistic modes. For example, Berg (1996, 2005) showed that humans display more sub-morphological non-contextual errors (errors not based on the syllabic or morphemic context in which it was found) in slips of the pen than in slips of the tongue, although the main form of errors were still predominantly contextual. In contrast, Mayberry (1995) showed that sign language users may have different underlying representations based on their skill level. Expert signers made predominantly conceptual errors (categorical substitutions, etc.) and the novice signers made predominantly articulatory errors (errors based on the shape and position of the hands), suggesting that the initial representation of a new language may be based on surface structure and only once the language is fully integrated do the conceptual representations dominate.

Looking at errors in nonhumans, Gardner and Gardner (1984) reported that sign-language-using chimpanzees made errors that were predominantly articulatory, similar to the error pattern that Mayberry (1995) associated with novice level signing skill. However, the chimpanzees did make errors within the categories of Animates, Food, and Drinks more often than chance (Gardner and Gardner 1984) a pattern indicating representation on the level of semantic categories. For example, Washoe signed CAT for dog and FLOWER TREE LEAF FLOWER for a picture of daisies.

Categorization in humans and nonhumans

Human categories are known to consist of multi-layered, integrated webs of associations and those associations, typically culturally based, define human categories (Nelson 1974). Most researchers agree that infants begin by categorizing objects perceptually and only later in their first year do they begin to form conceptual categories, an idea first proposed by Nelson (1974) and since confirmed by a number of studies (Bovet et al. 2005; Tomikawa and Dodd 1980). For example, Mandler and McDonough (1993) found, utilizing the familiarization paradigm, that infants of around 7 months of age did not reliably distinguish between pictures of airplanes and pictures of birds, due to the perceptual similarity. Infants of 11 months did make this distinction based on the conceptual difference (animals vs. machines).

An additional debate concerns the hierarchical level of categorization that is primary in human representation. "Basic" level labels (based on perceptual differences, and the least abstract level of categorization) are among the first nouns learned by human infants (Bloom 2000; Mervis 1987) and Rosch (1973) suggested that the initial categorical conceptions are also made at this "basic level". This theory has been refuted in a study by Mandler and Bauer (1988) who showed that children make initial categorizations at the most abstract level (a superordinate level) and only later break those categories down and arrive at basic-level categorization. The developmental sequence of global-basic categorization in humans has been further confused in later studies, with the exact categories formed dependant on stimuli received and methodology of the tests (for a review of relevant studies, see Mareschal and Quinn 2001).

In nonhumans, categorization is most frequently studied using operant conditioning techniques with same-different discriminations or delayed matching to sample paradigms. Researchers using these paradigms have shown that pigeons, baboons and other animals can form categories (e.g. Bovet and Vauclair 1998; Quinn et al. 2001; Young and Wasserman 2001) and at several levels of abstraction (e.g. Lazareva et al. 2004; Vonk and MacDonald 2002). However, a recent study confirmed that 3-year-old children do not require this training to pass a comparable task (Bovet et al. 2005).

Similarly, recent work has shown that chimpanzees can categorize natural basic-level categories without specific training (Brown and Boysen 2000; Murai et al. 2005; Tanaka 2001) and that gorillas and orangutans can form abstract categories and utilize second order abstractions to categorize shapes as same or different (Vonk 2003; Vonk and MacDonald 2004). However, these studies have not, with two exceptions (Vonk 2002, 2003) eliminated the possibility of success based solely on perceptual categorization. For example, Vonk (2003) trained a gorilla and four orangutans to perform a delayed matching to sample task based on simple shape and color. Later testing revealed that the gorilla and three orangutans could complete the task based on the relationship between two items (same shape or color—second level abstraction), although these relationships were still based on a perceptual similarity. Similar abilities have been shown in an African Grey parrot, Alex, who can correctly label what is the same or different about two items (color, shape, or material) (Pepperberg 1987) and can label the correct number of items in a recursive task that requires several levels of categorization (Pepperberg 1992).

Symbolic abilities in two bonobos

Savage-Rumbaugh et al. (1986, 1993) have explored the symbolic capacities of bonobos (*Pan paniscus*) and chimpanzees

(*Pan troglodytes*) in studies that have indicated that, when reared in an environment with English, a keyboard of visual symbols (lexigrams), and cultural emersion, apes can acquire symbolic capacities without explicit training (Brakke and Savage-Rumbaugh 1995, 1996; Greenfield and Savage-Rumbaugh 1991; Lyn and Savage-Rumbaugh 2000; Savage-Rumbaugh et al. 1980, 1986, 1993). Among others, two bonobos (*Pan paniscus*), Kanzi—a male born in 1980, and Panbanisha—a female born in 1985, have demonstrated the abilities to use lexigrams to name objects in double blind studies (Brakke and Savage-Rumbaugh 1995, 1996; Savage-Rumbaugh et al. 1986), to associate novel English names with novel objects with very few exposures to both object and word (Lyn and Savage-Rumbaugh 2000), to utilize imitation in an intentionally communicative context (Greenfield 1980), to make semantically-based combinations across both lexigram and gestural combinations (Greenfield and Savage-Rumbaugh 1990, 1991; Greenfield and Lyn 2007, also Lyn et al., unpublished data), and to comprehend English sentences at least at a similar level to a two-and-a-half year old child tested in the same manner as the ape (Savage-Rumbaugh et al. 1993).

The present study utilizes error analysis techniques from child language research designed to study linguistic representation in children to study representations of a human-devised symbolic system in a closely related species (the bonobo). The present study utilizes data gleaned over many years of vocabulary testing to address the question of the mental organization of the bonobos' lexigram vocabulary and its associations with English words and referents. Because these errors were never specifically trained nor were the bonobos specifically rewarded for lexigram use during acquisition, any regularity found would reveal the ape's spontaneous mental organization rather than human shaping.

Methods and materials

Subjects and rearing environment

Two bonobos (*Pan paniscus*) reared and housed at the Language Research Center in Atlanta, Georgia served as subjects. Kanzi—a male born in 1980, and Panbanisha—a female born in 1985 are half-siblings and were reared in an environment utilizing within-species and cross-species communication, in the form of gesture, speech, and written visual symbols (lexigrams) placed on a keyboard (Brakke and Savage-Rumbaugh 1995, 1996; Savage-Rumbaugh 1986; Savage-Rumbaugh et al. 1986, 1993). This keyboard was printed with up to 384 visuographic symbols (lexigrams) that were designed to be non-iconic and were randomly arranged. Therefore, neither the appearance of

the lexigram nor the position of the lexigram on the keyboard gave any clues as to the meaning. This differs from earlier ape language research using lexigram symbols in which the physical appearance of the lexigram was associated with referential or syntactical information (Rumbaugh 1977). For the majority of this study (8 of the 10 years), the keyboard included 256 symbols (See S1 for more details).

During the acquisition of lexigram use, the apes began to comprehend English and utilize the lexigram keyboard observationally, without specific teaching methods such as operant training techniques or food/specific reinforcement as had been used in other ape language studies (Gardner and Gardner 1969; Rumbaugh 1977; Savage-Rumbaugh 1986; Terrace 1979), although social interaction similar to that given to children acquiring language was a part of the ape/caretaker interaction. Food rewards were not given for English comprehension nor for keyboard use, rather symbol learning was accomplished in a social environment where English and the keyboard were used to announce movements, plans, activities and objects. Caregivers and other apes (some symbol-competent apes and some not) were in the rearing environment with the apes 24 h/day 7 days/week and utilized both English and the keyboard during all waking hours.

Daily activities for the apes at the Language Research Center included the use of lexigrams and gestures to request movement throughout the 50-acre lab property including the woods, living areas, and entrance into any of the animal cages as well as the use of gestures and lexigrams to comment on or make statements about objects or actions in their environs. Locations both inside and outside were given labels and specific foods were associated with those locations (for example, bananas were found at the treehouse, whereas juice was found at A-frame). Animals such as dogs, wildlife, and costumed characters such as people in bunny suits and gorilla suits were also part of the daily routine. These characters would act out scripted interactions, but could also act out actions specified by the apes (for further examples of interactions and earlier findings from these studies, see Brakke and Savage-Rumbaugh 1995, 1996; Lyn and Savage-Rumbaugh 2000; Savage-Rumbaugh 1986; Savage-Rumbaugh et al. 1986, 1993).

Testing procedures

Between the years 1990 and 2001, the English and lexigram vocabularies of Panbanisha and Kanzi were tested on a regular basis. The paradigmatic vocabulary test was to present a sample in English and require the subject to choose the corresponding lexigram from their keyboard of 256 or more symbols, similar to human tests of reading vocabulary in which the experimenter says a word and the

subject is to find that word in a printed list. Although the apes were not operantly rewarded for correct answers or punished for incorrect answers, the caregivers could respond by communicating the correctness of the answer to the apes (e.g. “yes, that’s right” Or “No, find the blueberries”), therefore the apes were aware that there was a “correct” answer and “erroneous” answers. These responses could be considered a verbal reward, although it is important to note that these were tests to measure the comprehension that the apes had already acquired, not the mechanism by which the apes acquired symbols.

Alternative tests included as part of the data set could present, as the sample, either an English word, a lexigram, a picture, or any combination of English word, lexigram, and picture, with the response still being on the keyboard (no test in which the alternatives were chosen from a smaller set than 256 alternatives was included in the analysis). All tests were run according to the procedure described above, with only the sample type differing, however, not all tests were double-blind, neither were double blind-tests necessarily differentiated. Therefore, all trials should be assumed to be non-double-blind. Tests including all sample types are included in the analysis to allow exploration into all representations revealed by the apes’ associations between English word, lexigram symbol, and physical referent (represented in the tests as a photograph). These vocabulary tests and other tests that presented full sentences yielded volumes of published data, indicating that the apes could match English words, photographs and lexigrams to each other (Brakke and Savage-Rumbaugh 1995, 1996; Savage-Rumbaugh et al. 1986).

Data set

The error analysis was conducted on vocabulary tests recorded at the Language Research Center between 1990 and 2001 comprised of 7,563 trials administered to Panbanisha and 6,091 trials to Kanzi. Overall, the bonobos’ error rates across all trials were relatively low—Panbanisha: 5.7% (88.9% correct, 5.4% correct with assistance or refusal to answer); Kanzi: 17.5% (77.7% correct, 4.8% assisted or refused); resulting in a data pool of 427 errors for Panbanisha and 1,070 errors for Kanzi. Please note that the chance rate for a correct response for the majority of the study was 1/256 (and for the final 3 years chance was 1/384)—the denominators representing the number of lexigrams on the lexigram boards. Importantly, these results show fundamental comprehension of these symbols by the apes at the beginning of testing.

Errors were coded whenever an ape indicated the incorrect lexigram; this may be an overestimate of answers that were “errors” in the minds of the apes. Many of these “errors” may have been attempted communications rather

than attempts to respond to the sample, such as requests for items, continuations of previous topics, etc. and are possibilities for further studies. However, the low error rates suggest that the apes were familiar with the task and were aware of the existence of correct and incorrect answers.

Additionally, the apes frequently self-corrected when an error was made (self-correction was noted by the individual experimenters when the ape selected the correct lexigram after an error was made, but before any feedback was given by the experimenter). Kanzi self corrected on 89/1,070 errors (8%) and Panbanisha on 49/428 errors (11%), indicating acknowledgement of errors. Importantly, our results should only be affected negatively by these extraneous “errors”. Any communicative attempts that are not related to the sample (and therefore the mental representations associated with the sample) should bear no regular resemblance to the sample, therefore increasing the rate of uncatégorizable errors. However, to clarify any differences between these self-corrections and other errors, a brief analysis of these errors alone is presented below.


Coding

Errors were coded according to three dimensions of similarity between sample and erroneous choice: auditory similarity, physical similarity of the referent, and physical similarity of the lexigrams (see Table 1). Errors were also coded for two semantic error types—categorical equivalence and spatio-temporal associations; one syntactic error type—part of speech; and one articulatory error type—proximity on the keyboard. All remaining errors were classified as “uncategorizable”, although these errors occasionally included likely attempted communications (as opposed to responses to the sample). Each error was coded with all of the possible error types to which it matched; that is, errors could be placed into more than one coding category.

Reliability coding

An experimenter who had spent over 6 years working with the keyboard and the apes was employed as a reliability coder. Due to the specific use of the second coder to check reliability, all disagreements were resolved in favor of the first author. The reliability coder was first given a total of 175 training trials and then coded 174 new trials that were compared to the same trials coded by the first author. The first author coded the trials with 307 separate codes, the reliability coder with 311. The two experimenters agreed on 263 of these codes and disagreed 66 times, for a basic agreement of 80%. Because of the low distribution of some code types, inter-observer reliability for these codes was determined using Cohen’s Kappa (see Table 1 for Kappa values). Agreement ranges from a high of 0.94 for proxim-

Table 1 Error type codes: error types, definitions and Cohen's Kappa for interobserver reliability

Error types	Description	Examples	Cohen's Kappa
Auditory similarity	The sample and the error have similar spoken labels	Jared/cherries Jelly/Jell-O	0.94
Physical similarity	A photograph of the sample could be mistaken for a photograph of the error	Apples/apricots River/water	0.66
Lexigram similarity	The lexigrams of the sample and the error have at least two physical similarities (e.g. color, shape of design element, number of design elements)		0.76
Categorical equivalence	The sample and the error belong to a common category (e.g. location, edibles, animates)	Lemonade/juice Sue's office/A-frame Towel/blanket	0.86
Spatio-temporal association	The sample and the error are frequently found together in time or space (and/or are functionally related), but are not categorical equivalents (e.g. a location and the food found there; an object and the instrument to open it)	Key/outside Criss cross corners/blackberries	0.61
Part of speech	The sample and the error belong to the same syntactical category, but are not categorical equivalents or spatio-temporal associations (e.g. nouns)	Sue/milk Go/tickle	0.62
Proximity	On the keyboard, the lexigram of the error is within 4 lexigrams in any direction of the lexigram on the sample (e.g. one above the other on the keyboard)	sweet potato/chase	0.94
Uncategorizable	The error cannot be placed in to any of the above categories		0.73

All Kappa values are in the good to excellent range

ity to a low of 0.61 for spatio-temporal association. All are within acceptable ranges. Lower scores are in part due to the coding scheme—e.g. spatio-temporal association is defined as “related in space and time, but not a categorical equivalent”. Therefore, any disagreement on categorical equivalence would also affect reliability on spatio-temporal association (and also part of speech).

Secondary coding

Our initial categorization coding represented high-level categories (e.g. edibles, locations)—more abstract than basic-level categories. To explore the apes understanding of basic-level categories, the data were recoded. Errors coded within the larger category “edibles”, were recoded to reflect basic level categories including: drinks, meat, grains, vegetables, fruit, dairy, sweet, spice, water, and wild foods

found in the woods. Errors coded within a “location” category were re-coded: indoor or outdoor. Errors coded as an “animate” were recoded: people, apes, animals, and costumed characters. Errors coded as “objects” were recoded: grooming, toys, fire, clothes, movement, tool, and objects found in the woods. Errors coded as “actions” were recoded: movement or play. Most categories also had superordinate examples such as “food” or “outdoors”, both lexigrams found on the keyboard.

These recodes were specifically determined from a list—for example: the indoor locations were childside, colony room, tool room, middle test room, bedroom, group room, colony room, kitchen, bathroom, t-room, staff office, playyard, and Sue's office; and the list of “animals” included dogs, bugs, and snakes, but excluded apes and humans. A very few items were included on more than one list. For instance, Kool-Aid appeared on both the “sweet” list and

the “drink” list. In these cases, all applicable sub-categorical matches were coded [Kool-Aid was a categorical match for both Jello (a sweet) and orange juice (a drink)]. These lists pre-supposed certain categories (for example, “play-yard” could have been categorized as outdoors) and further analysis may reveal that the apes did not conform to all of these pre-suppositions.

Further coding of the “proximity” variable was also undertaken to investigate the apes’ representation of key location. 2-lexigram (and 1-lexigram) proximity were coded when the error was within 2 (or 1) lexigrams of the sample on the keyboard.

Analysis

Random distributions

A good measure of chance based solely on the lexigram keyboard was difficult to obtain because each lexigram would have its own probability of being coded with a specific error type. For example, “coke” could be a categorical equivalent to 63 other “edibles” on the keyboard. However, a location such as “colony room” has only 32 possibilities for categorical equivalence. Also, lexigrams in the middle of the keyboard could be coded as “proximity” to 80 other lexigrams, whereas those in the corners could only be coded with 24 other lexigrams.

Therefore, in order to separate meaningful error patterns from chance patterns, a random distribution of 300 sample and error pairs was created to compare actual errors to predicted errors, based on each lexigram being equally likely to serve as a sample or an error. Microsoft Excel’s random number generator yielded an integer from a uniform distribution of the numbers between 1 and 256 for each sample and error, which were mapped one-to-one to the lexigrams and spoken English glosses on the 256-lexigram keyboard and were coded in the same fashion as the apes’ responses. The smaller keyboard was utilized to provide a conservative estimate of chance.

For example, the random number distribution would yield numbers between 1 and 300 with each number being equally likely. An example of a random pair would be 104 and 25. 104 was glossed as “balloon” and 25 as “celery”. This would be considered a trial in which “balloon” was presented and “celery” was the error made, which was then coded with “part of speech”. By this method, individual probabilities of specific lexigrams should be as likely in the random distribution as in the apes’ errors.

A second random distribution was also analyzed to explore possible effects of the caretakers’ or the apes’ preferences for specific lexigrams as samples and/or errors. Because not all of the sample items were chosen randomly, preferences for particular lexigrams could produce spurious results. For this analysis, two random distributions were

generated by calculating the percentage of the apes’ data containing a specific referent, first for samples, then for errors. Using the discrete distributions created by these percentages, 300 random numbers were generated, again, first for samples and then for errors. Excel’s discrete distribution random number generator was used to perform this task. Because Excel’s random number generator initiates with a 1, an additional random number (total of 301) was generated for the errors and the first number (1) was discarded.

Sample modes

As described above, the vocabulary sample for the tests could include any combination of English, lexigram and photographic representation. A breakdown of error types when a single sample type (English, Lexigram, or Photograph) is presented is analyzed below. Also, a correlation analysis investigated the effects of the presence of the three sample modes on the types of errors made to determine the amount of variance in the error type that could be accounted for by sample mode. Note that in these correlations, a trial with English + Lexigram as a sample would be included in both the correlation for English sample type and the correlation for Lexigram sample type. This is in contrast to the breakdown of error types by sample type, which data set was restricted to samples that only included one sample type.

The hypotheses being tested in these analyses predicted that, for example, if the sample includes the sample type “photograph”, the apes’ visual representation of the symbol should be activated and more errors of the type “physical similarity” should be reported. A specific example might be that a sample includes a picture of an apricot. The ape sees the apricot and its visual representation of the picture includes “round” and “orange”, this would make an error of “peach” or “orange” more likely. This hypothesis is most revealing when the sample includes a lexigram. Since the act of choosing a lexigram on the keyboard is more similar to a multiple-choice task than to a recall task (which is what sign-language or other language tests are), the ape should be able to make a purely visual match of lexigram to lexigram without any other representations being triggered—therefore lexigram similarity errors should be the most frequent and, in a correlation between lexigram sample and lexigram similarity errors, the sample type (lexigram) should make a large contribution to the variance.

Results

General results

The error type results for the apes and the random distribution appear in Table 2 (See S2 for all results and codes).

Table 2 Error type results: number of errors (and percentage of total errors) coded in each error type for each ape and both random distributions (rd uniform and rd matched)

Error types	Kanzi	Panbanisha	rd uniform	rd matched
Auditory similarity	97* (9%)	73* (17%)	2 (0.6%)	3 (1%)
Physical similarity	130* (12%)	53* (12%)	6 (3%)	4 (1%)
Lexigram similarity	113** (11%)	66* (15%)	14 (5%)	9 (3%)
Categorical equivalence	607* (57%)	211* (49%)	46 (15%)	47 (16%)
Spatio-temporal association	68 + (6%)	64* (28%)	4 (1%)	8 (3%)
Part of speech	213^ (20%)	120^ (28%)	131 (43%)	144 (48%)
Proximity	381* (36%)	239* (56%)	76 (25%)	48 (16%)
Uncategorizable	140^ (13%)	45^ (11%)	86 (29%)	75 (25%)
Total errors	1070	427	300	300

Errors can be coded with more than one error type (χ^2 tests)

* $P < 0.001$, apes made more errors of this type than would be expected by either rd

** $P < 0.006$, apes made more errors of this type than would be expected by either rd

^ $P < 0.001$, apes made fewer errors of this type than would be expected by either rd;

+ Kanzi made more errors of this type than would be expected— $P < 0.05$ compared to rd matched, $P < 0.001$ compared to rd uniform

Uncategorizable errors were recorded in the apes' responses significantly less often than would be expected by chance as measured against both random distributions. Errors were associated more often than predicted by chance with auditory similarity, physical similarity, lexigram similarity, proximity, categorical equivalence, and spatio-temporal association; but less often than would be predicted by chance with "part of speech" (see Fig. 1 for example). Because each of our hypotheses is being tested independently of the others (Does auditory similarity effect error type? Does categorical equivalence effect error type?), Bonferroni's adjustment is not particularly called for (see Perneger 1998). However, should a Bonferroni's adjustment be utilized, it would require a P value of under 0.006 for each comparison for a total alpha value of $P < 0.05$ for all comparisons between each ape and each random distribution. Only one of our comparisons does not meet this P value, that of spatio-temporal association between Kanzi and the matched random distribution (rd matched, $\chi^2(1, 1446) = 5.55$, $P = 0.02$). Finally, errors were more frequently coded with more than one code than would be expected by either random distribution (rd) (Kanzi 478/1070 (45%); Panbanisha 278/427 (65%); rd uniform 41/300 (14%); rd matched 39/300 (13%); $P < 0.001$, χ^2 tests).

Self-corrected errors

To further explore self-corrected errors, a total of 89 errors by Kanzi and 49 errors by Panbanisha were compared to the two random distributions. Both Kanzi and Panbanisha made fewer uncategorizable errors when they self-corrected than when they did not, although this difference was only significant for Kanzi (Kanzi: 4/89 (4%), $\chi^2(1, 1303) = 4.64$,

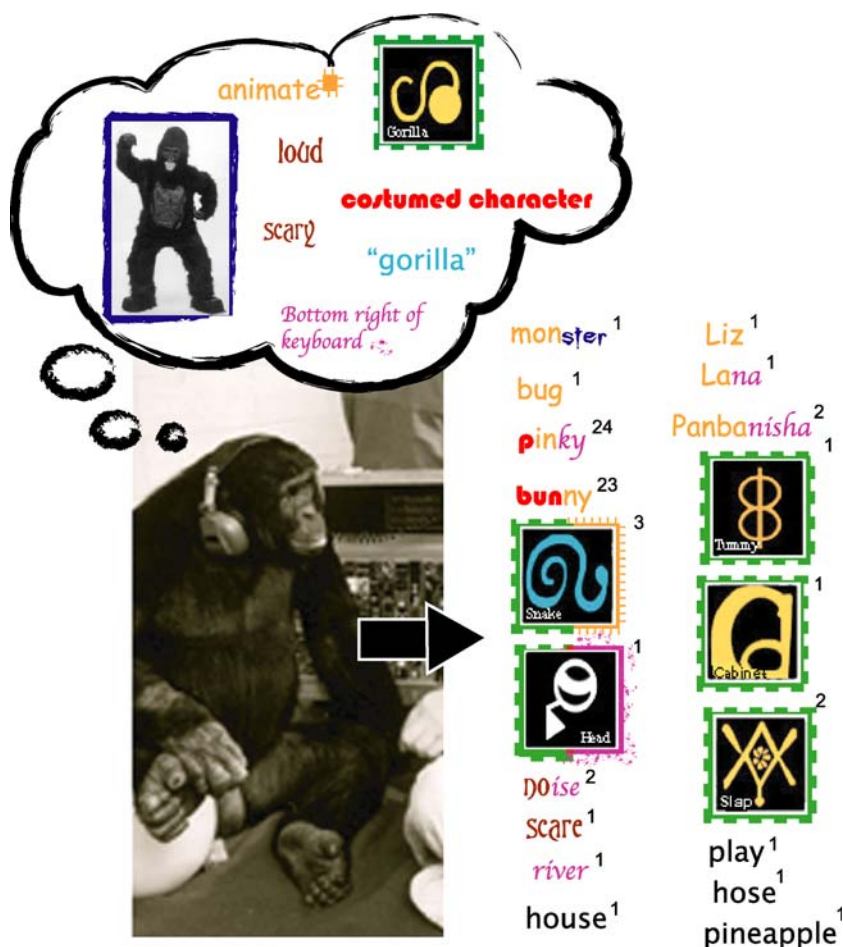
$P = 0.03$; Panbanisha: 2/49 (4%), $\chi^2(1, 523) = 1.77$, $P = 0.18$). All of Panbanisha's error types followed the pattern of her general errors (when all errors were included)—more errors than would be expected in auditory similarity (5/49, 10%), physical similarity (4/49, 8%), lexigram similarity (10/49, 20%), proximity (36/49, 73%), categorical equivalence (19/49, 39%), and spatio-temporal association (16/49, 33%); but less than would be predicted by chance with "part of speech" (7/49, 14%). Comparing her self-corrected results with her general errors resulted in no differences except for spatio-temporal association, in which she made more errors when she self-corrected ($\chi^2(1, 556) = 6.24$, $P = 0.01$).

In contrast, Kanzi's error types followed the pattern of his general errors in physical similarity (11/89, 12%), lexigram similarity (20/89, 22%), proximity (40/89, 45%), categorical equivalence (67/89, 75%) (more errors than in either rd) and "part of speech" (12/89, 13%) (fewer errors than would be expected by either rd); however his results in auditory similarity (1/89, 1%) and spatio-temporal association (1/89, 1%) did not differ from chance. Comparing his self-corrected errors to his general errors, Kanzi made fewer auditory similarity errors when he self-corrected (1/89, 1%; $\chi^2(1, 1257) = 6.02$, $P = 0.01$) and more lexigram similarity errors (20/89, 22%; $\chi^2(1, 1292) = 8.36$, $P = 0.003$); however, his rate of self-corrected spatio-temporal errors were not significantly different from his general errors, possibly due to low numbers (1/89, 1%; $\chi^2(1, 1228) = 3.72$, $P = 0.05$).

Sample mode

Panbanisha and Kanzi's distribution of error types differed depending on the mode of the sample (English, lexigram, or photograph) (Panbanisha $\chi^2(1, 745) = 18.14$, $P < 0.001$;

Fig. 1 Associations suggested by the data to exist in the minds of the apes. Representation of the semantic associations in the minds of the apes (*left*) and the errors (*right*) that were generated in connection to the symbol for “gorilla”. Errors are marked with colors, typefaces and line types to match the error type codes and the number following the error indicates the number of times those symbols were confused in this study



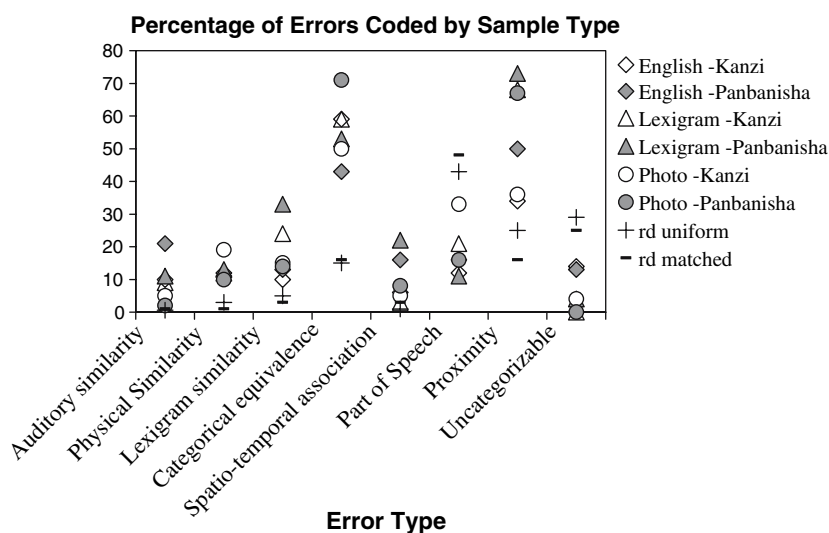
Kanzi $\chi^2(1, 1551) > 30, P < 0.001$ (Fig. 2), however, the comparisons between the apes' results and the random distributions did not differ greatly. For Kanzi, only two (out of 16) results that were significant with all errors included were not significantly different from either random distribution when separated by sample type: spatio-temporal association with a lexigram sample and part of speech with a photograph sample. Three other of Kanzi's results were not significantly different from one random distribution: part of speech with a lexigram sample and proximity with a photograph sample (not significantly different from rd uniform); and spatio-temporal association with a photograph sample (not significantly different from rd matched). For Panbanisha, one result was not significantly different from either random distribution: auditory similarity with a photograph sample; and another was not significantly different from one random distribution: spatio-temporal association with a photograph sample (not significantly different from rd matched). Many of these drops in significance can be accorded to low numbers (for example, there were only 51 examples of errors for Panbanisha with a photograph sample).

Importantly, the majority of findings are unaffected by sample type. For both Kanzi and Panbanisha, auditory similarity

errors were more likely than would be expected by chance even when the sample was a lexigram (and for Kanzi, more likely than chance even when the sample was a photograph). Similarly, physical similarity errors were found more often than would be expected by chance for both apes when the sample was only the English word or only a lexigram. For both apes, categorical equivalence and proximity were the most likely errors to be found, regardless of sample type. For example, when the sample was a lexigram and the apes therefore could have simply made a visual one-to-one match between the sample and the lexigram on the keyboard, many errors similar to the following were reported: Panbanisha: out-doors—lookout point (code: categorical equivalent), Kanzi: Sue's office—Sue (codes: auditory similarity, part of speech).

The correlations between the presence of a sample mode and error type also reveal no global effects. Panbanisha's rate of Auditory Similarity errors was correlated with samples that included English (63/341 errors from a sample that included English coded as auditory similarity; $r = 0.16, P < 0.001$) and her rate of lexigram similarity errors was correlated with sample that included a lexigram (15/46; $r = 0.17, P < 0.001$), but her rate of physical similarity errors was not associated with a sample that included a

Fig. 2 Percentages of error type codes for each ape and each random distribution (*rd*) when the results are separated by sample type. *Note* this figure only includes errors when the sample type were only English, only Lexigram, or only Photograph, therefore, the entire database, which includes a small portion of samples of English + Photo, English + Lexigram, and Photo + Lexigram, is not represented



photograph; Kanzi, in contrast, showed a significant increase in physical similarity errors when the sample contained a photograph (23/128; $r = 0.08$, $P < 0.01$), but showed no effect on auditory similarity by samples that included English and no effect on lexigram similarity errors by samples that included a lexigram.

Correlations allowed us not only a measure of significance, but of the variance associated with a specific independent variable (in this case, sample type), when other variables are accounted for (in this case, the other variables that may have been correlated with the apes' rate of specific error types included: year of the test, specific sample (some lexigrams may have been more readily confused than others), other codes (for example, proximity errors also increased when a sample included a lexigram, Fig. 2), etc). While these correlations are statistically significant, the r^2 s, and therefore the variance associated with these sample types, are quite small (largest is Panbanisha's r^2 between lexigram samples and lexigram similarity errors -0.0289 —whether the sample contains a lexigram accounts for less than 3% of the variance associated with Panbanisha making an error based on lexigram similarity).

Error types

Proximity

The apes showed positive results for proximity, that is, they made erroneous choices within 4 lexigrams of the correct choice more often than would be expected by chance as measured by either random distribution. Because this measure is unlikely to specifically identify articulatory errors, as more than 1/2 of one panel of the keyboard would be considered in “proximity” to a lexigram

in the center of the keyboard, further investigation was warranted. Recoding the proximal symbols showed that both apes also more frequently made errors within 2 lexigrams of the correct choice than either random distribution [Kanzi 202 errors (19%); Panbanisha 151 errors (35%); rd uniform 30 errors (10%); rd matched 16 errors (5%): all $P < 0.01$, χ^2 tests] and within 1 lexigram of the correct choice (Kanzi 95 errors (9%); Panbanisha 72 errors (17%); rd uniform 14 errors (5%); rd matched 5 errors (2%): all $P < 0.05$). When comparing the differences between 2-lexigram proximity and 4-lexigram proximity (in other words, the errors made between 3 and 4 lexigrams away from the correct choice) both Kanzi (179/381 proximity errors, 47%) and Panbanisha (88/239, 37%) made significantly fewer errors than both random distributions [rd uniform (46/76, 61%); rd matched 32/48 (67%), all $P < 0.05$, χ^2 tests]. This pattern was not the case for the differences between 2-lexigram and 1-lexigram proximity [Kanzi 107/202 (53%); Panbanisha 79/151 (52%); rd uniform 16/30 (53%); rd matched 11/16 (69%), all $P > 0.05$, χ^2 tests], suggesting that the most salient level of proximity for the apes may be within 2 lexigrams and that proximity may be an initial search mechanism for the apes, with the final choice of a lexigram also guided by other mechanisms (see later analyses).

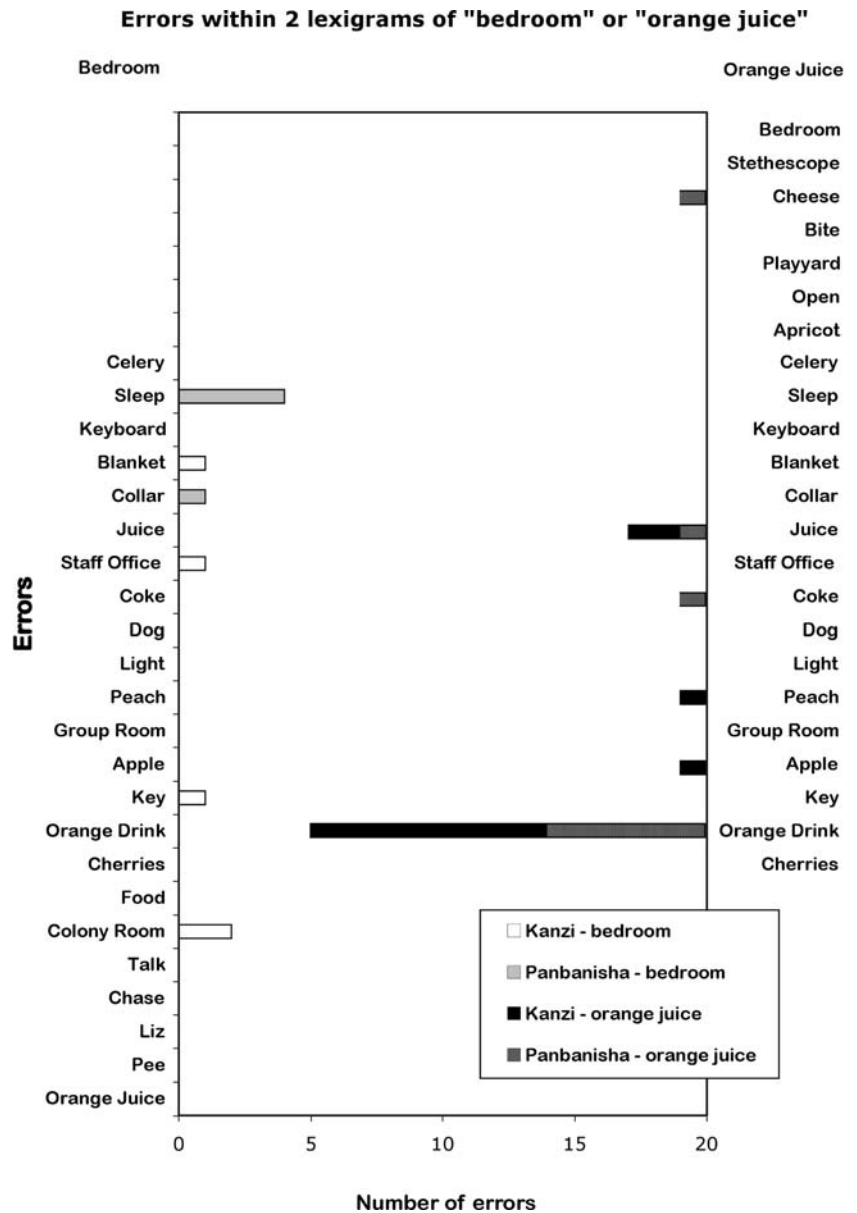
To insure that proximity errors did not drive all results, a brief analysis was performed in which all proximity errors were removed from all databases. Removal of all proximity errors did not result in substantially different results. Both apes had error rates that were significantly higher than chance in auditory similarity [Kanzi 75/745 (10%), Panbanisha 51/247 (21%), rd uniform 2/224 (1%), rd matched 2/265 (1%)], physical similarity [Kanzi 77/745 (10%), Panbanisha 19/247 (8%), rd uniform 5/224 (2%), rd matched 3/265 (1%)], lexigram similarity [Kanzi 56/745

(8%), Panbanisha 22/247 (9%), rd uniform 6/224 (3%), rd matched 9/265 (3%), and categorical equivalence [Kanzi 383/745 (51%), Panbanisha 51/247 (21%), rd uniform 34/224 (15%), rd matched 36/265 (14%)]. Similarly, both apes had error rates that were less likely than chance in part of speech [Kanzi 99/745 (13%), Panbanisha 30/247 (12%), rd uniform 136/224 (61%), rd matched 129/265 (49%)] and uncategorizable errors [Kanzi 140/745 (19%), Panbanisha 44/247 (18%), rd uniform 86/224 (38%), rd matched 75/265 (28%), all χ^2 tests]. Only the error rate for spatio-temporal association showed no significant difference to a random distribution and then, only to the matched random distribution [Kanzi 35/745 (5%)—rd uniform 2/224 (1%):

$\chi^2(1, 1006) = 6.41, (P = 0.01$; rd matched 7/265 (3%) $\chi^2(1, 1052) = 1.92, (P = 0.16$; Panbanisha 13/247 (5%)—rd uniform $\chi^2(1, 486) = 6.84, (P = 0.009$; rd matched $\chi^2(1, 532) = 2.16, P = 0.14$].

In addition, the apes made fewer errors that were coded only “proximity” or “proximity” combined only with “part of speech” (Kanzi = 94 out of 381 proximity errors (33%); Panbanisha = 38/239 (16%)) than would be expected by either random distribution (rd uniform 54/76 (71%); rd matched 35/48 (73%), $P < 0.001$, χ^2 tests). Instead, most erroneous choices were both proximal to the correct choice, and associated to it in another way (see Fig. 3 for an example of differential errors associated with nearby lexigrams).

Fig. 3 Proximity errors. Number of specific errors within 2-lexigram proximity made by Kanzi and Panbanisha in response to “bedroom” and “orange juice” samples. “Bedroom” and “orange juice” are side-by-side on the keyboard, therefore the majority of proximal lexigrams are shared, however, the errors made by the bonobos are mutually exclusive, depending on other associations with the sample



Part of speech

The apes were significantly less likely to make an error based on “part of speech” than would be expected by chance. Interestingly, both Kanzi and Panbanisha made more verb errors when presented with a verb than would be expected by chance alone (Kanzi 88/155 verb errors when sample was a verb; Panbanisha 14/41; both $P < 0.05$, binomial test, probability of choosing a verb by chance 9%). Note, however, that even though both apes chose verbs more frequently when a verb was the sample, 21 out of the 23 verbs on the keyboard are actions, so semantic class cannot be separated from syntactical class in this example.

Categorical errors

Some of the apes’ categorical errors overlapped with similarity errors [ex: perceptual categorization, e.g. cherries and blueberries (foods), lighter and sparkler (objects), Sherman and Austin (animates), Scrubby Pine Nook and Mushroom Trail (locations)]. However, even when this group of responses was removed from the overall analysis (therefore removing all categorizations with perceptual similarity), the remaining categorical errors were found significantly more often than chance [conceptual categorical errors—Kanzi 488 (45.6% of all his errors); Panbanisha 163 (38%); random distributions: rd uniform 42 (14.0%); rd matched 44 (15%) ($P < 0.001$, χ^2 tests, for both Kanzi and Panbanisha)]. See Fig. 4 for one example. Other examples of conceptual categorization include: peaches and blueberries (foods), balloon and sparkler (objects), Austin and dog (animates), hilltop and river (locations).

When the categorical errors were re-analyzed for basic-level and higher-order categorization, Panbanisha had basic level matches on 161/211 categorical errors (76%) and Kanzi had 388/607 (64%) [both significantly higher than both random distributions—rd uniform 13/46 (28%), rd

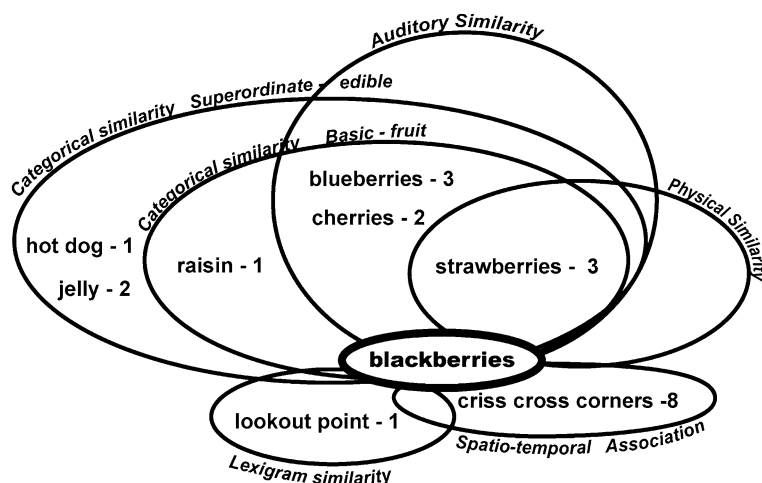
matched 20/47 (43%) ($P < 0.01$, χ^2 tests)]. Kanzi made 219 errors at the higher level of categorization (20.5% of all Kanzi’s errors), significantly more than either random distribution—rd uniform 33 (11%); rd matched 27 (9%) (rd uniform $\chi^2(1, 1370) = 13.91$, $P < 0.001$; rd matched $\chi^2(1, 1370) = 20.91$, $P < 0.001$). In contrast, Panbanisha’s errors were predominantly basic-level matches with only 50 higher-level matches (12% of all errors) (rd uniform $\chi^2(1, 727) = 0.09$, $P = 0.77$; rd matched $\chi^2(1, 727) = 1.37$, $P = 0.24$).

Discussion

These findings support the suggestion that the symbolic mental representations formed by nonhuman apes can spontaneously organize into integrated, hierarchical categorical representational systems. This claim is supported by the fact that the apes’ errors were more likely than expected by chance to be coded with more than one error type and these error types (except for part of speech) were all found more often than would be expected by chance. These findings argue against a simple associative mechanism for the production of these errors. More likely, the representations interact to produce the error patterns seen in these data, forming integrated representational systems. For example, both Kanzi and Panbanisha made fewer errors that were coded as only “proximity” or “proximity” and “part of speech” than would be expected by either random distribution (Fig. 3). This may suggest that an initial search mechanism in response to a sample is the apes’ representation of the lexigram’s location on the board, but the selection of an error is finally guided by other representations, such as categorical equivalence or spatio-temporal association.

The difference in the apes’ results depending on the sample type presented may reflect different underlying representational systems depending on the trigger mechanism.

Fig. 4 Conceptual and hierarchical categorization. Representation of the possible error types (as coded in this study) and the number of errors made by the apes within those error types connected with one example, the symbol “blackberries”. The more circles overlap, the more categories of similarity between the error and the sample, and also the more likely the confusion between the two in these tests



Similarly, in humans, error-producing mechanisms (and possibly the underlying representational systems) seem to differ by production mode (Berg 1996, 2005). However, the low variance associated with sample type and the fundamental similarities between the percentages of error types coded when broken down by sample type suggest that these representational systems are not completely separate, nor are they solely triggered by physical stimulus, but rather, are triggered by all stimuli that evoke the representation.

Hierarchical representation can be argued based on the apes' categorical error types. Within these errors, Panbanisha and Kanzi's high number of basic-level category matches suggest that this is the most salient categorization level for these bonobos as it is for humans. Higher-level categorization is suggested, however, by Kanzi's error pattern as it included more errors than were expected by the random distribution at both the basic level (when the higher-level matches were removed from the analysis) and the superordinate level (when all basic level matches were removed). These data would not be found if Kanzi only had basic-level representation (such as fruit, e.g. apple and orange) as there would be fewer than expected higher-level matches that did not match at the basic level (e.g. hot dog and orange), and *visa versa* for higher-level representation. These data demonstrate that Kanzi's mental representation of symbols not only contains categorical knowledge, but knowledge of several levels of categorization (see Fig. 4). The capacity for hierarchical categorization in nonhumans is confirmed by recent work with pigeons (Lazareva et al. 2004) and studies of gorillas and orangutans (Vonk and MacDonald 2002, 2004; Vonk 2003).

It is important to note that several levels of hierarchy are evident in the data from both Kanzi and Panbanisha. The most basic level of categorization [defined by Bloom (2000), pp. 148 as "the most inclusive level in which objects are judged to have many features in common"], is represented by many lexigrams themselves. This is the same level that has been frequently examined in recent non-human categorization studies including Brown and Boysen (2000) and includes such familiar and natural categories as dogs, noodles, oranges, and trees. Therefore, the "basic" level representation exhibited by Panbanisha, actually encompasses true basic-level categorization, e.g. "oranges" and a secondary level, "fruit", and Kanzi's representations include a third level, "edible".

This is not to say that the representations of the apes necessarily include this unifying concept explicitly. These data only show that the apes respond differentially to members of certain categories that were pre-defined by the experimenters than to non-members of these same categories. This is also true of most infant categorization studies (e.g. Levy and Schlesinger 1988; Madole and Oakes 1999; Markman 1989; Rosch 1988). Interestingly, some lexigrams

are specific, rather than categorical, such as Sue, Observation Room, and Kanzi and others are superordinates, such as food, outdoors, and a recent addition "things". The specifics of the apes' categorical constructions are good candidates for further research and analysis.

Our results also indicate that the apes make more conceptual categorical errors than would be expected by chance, paralleling findings from developmentally mature humans (Nelson 1977; Quinn et al. 2001). Categorical studies with pigeons, monkeys and 3-month old children (e.g. Bovet and Vauclair 1998; Brown and Boysen 2000; Quinn et al. 2001; Roberts and Mazmanian 1988; Tanaka 2001; Young and Wasserman 2001) have found perceptual categorization to be common, however, few studies of non-humans have shown conceptual categorization, and none of these were spontaneous (Bovet and Vauclair 1998; Savage-Rumbaugh et al. 1980; Young and Wasserman 2001). These results and earlier examples in Gardner and Gardner (1984) as well as the recent studies by Vonk (2002, 2003) support the suggestion that apes can form conceptual categories and that these categories are a part of these apes' representational systems.

Apart from these categorical representations, the results also reveal that the mental representations of the apes include spatio-temporal associations of symbolic information. Similarly, children and adults make a variety of spatio-temporal errors and word associations (Moran 1973; Nelson 1977), e.g. dark-night, chair-sit, and lion-cage. The bonobos' errors are similar, e.g. key-open, umbrella-rain, and monster-scare. Another spatio-temporal association frequently found in the apes' errors is the association of familiar locations with the food found at those locations, e.g. criss cross corners-blackberries, river-blueberries.

Additionally, both Kanzi and Panbanisha make more errors based on "proximity" than would be expected by chance, indicating that both apes' mental representations of the keyboard include information about where the lexigram is located on the board. Indeed, Panbanisha made more proximity errors than any other kind of error. However, further investigation of 2-lexigram and 1-lexigram proximity revealed that the most salient level of proximity is probably within 2 lexigrams of the correct answer and Panbanisha made fewer 2-lexigram proximity errors (35%) than categorical errors (49%).

These proximity data suggest that Kanzi and Panbanisha's errors were not predominantly articulatory, as are those of novice human sign language users (Mayberry 1995) and the sign-language using chimpanzees in the study by Gardner and Gardner (1984). In both of these studies, errors related to the location and/or shape of the sign were found without accompanying conceptual relations. These findings further support the hypothesis that the representations of the apes' are not predominantly

structured around the physical properties of keyboard (as would be suggested by the “novice” patterns of errors in the Mayberry study), but around the conceptual meanings and associations of those symbols (as would be suggested by the “expert” pattern of errors in the Mayberry study). Indeed, when proximity errors were removed from the analysis, almost all of the comparisons between the apes’ errors and those of the random distribution were maintained; the sole exception being spatio-temporal association (one possible explanation of this difference is that the representation of spatio-temporal association is most often triggered when the ape is searching the correct area of the keyboard for a particular symbol, but does not immediately find it).

Importantly, as not all of the vocabulary tests were run double-blind, any unconscious cuing on the part of the caretakers should have led to many proximity errors (including errors that had no other code), and this is not the pattern observed. These results further support the hypothesis that the apes’ representations of the lexigram symbols and their English and referential counterparts comprises a complex system of multiple associations, including physical, spatio-temporal and conceptual associations.

Many researchers have argued that syntactical structure is the singular linguistic ability that separates human language from animal communication (Calvin and Bickerton 2000; Pinker 1994; Wallman 1992). These data support that claim in that the apes did not make errors based on part of speech more often than would be expected by chance. However, the determination of chance is confounded because the “part of speech” code was only utilized in absence of a categorical match. Given the large number of categorical matches by the apes, and the small number of categorical matches in the random distributions, the detection of any syntactical markings would be difficult to separate from the semantic (categorical) effect. This confound is further affected by the predominance (over 86%) of nouns on the keyboard. Future analysis of data with more verb, adjective, and/or performative samples may clarify the syntactical marking question and would help to clarify the distinction between semantic relations (actions vs. agents, etc., see Greenfield and Savage-Rumbaugh 1990; Greenfield and Lyn 2007, also Lyn et al., unpublished data) and syntactical parts of speech (nouns vs. verbs, etc.). This distinction must be made, as although both apes make more verb errors when presented with a verb sample, the verb samples cannot be separated from their semantic class of “action”. The recently expanded keyboard contains many more nonaction verbs, allowing for possible future study.

Finally, there were a number of “error” types that included attempted communications, rather than attempts to respond to the sample, such as requests for items, continuations of previous topics, etc. For example, when asked to

find the lexigram for “Sue”, Kanzi initially responded with what could be interpreted as a request for interaction “Chase, Tickle” then indicated “Sue”. This interpretation is supported by the caretaker’s comment that Kanzi first responded with activities that he would like to do with Sue. In another instance, the caretaker’s comment at the beginning of a test session indicated that she and Kanzi had communicated about going out to the playyard. Kanzi responded “open” or “playyard” in all four errors found in that test session. Another example may be when an ape chooses the lexigram of a food found at a location, instead of the location itself (e.g. “blackberries” for “criss cross corners”). These “errors” are included within the “uncategorizable” code when none of the other above codes is indicated. However, these “errors” indicate the continued use of the keyboard as a communicative system and are an interesting subject for possible further analysis. Importantly, however, these future analyses would be reliant on the subjective views of caregivers present during the tests to tell us “why” the apes may have made the mistakes it did.

These findings detail the spontaneous formation of hierarchical, conceptual categories by symbol-using bonobos utilizing a data set of vocabulary errors. Many questions remain to be answered including questions of any developmental changes that may have taken place over the 10+ years of data collection, co-occurrences of certain kinds of errors; detailed examination of the uncategorizable errors, and analysis of the specific types of hierarchical categories that were formed by the bonobos. While outside of the scope of this paper, these questions remain worthy of further pursuit.

Acknowledgments The author wishes to acknowledge E. Sue Savage-Rumbaugh without whose determination, strength, and insight into the apes no work on Kanzi and Panbanisha’s abilities would be possible. Also, the author wishes to acknowledge the contributions of the staff of the Language Research Center, Atlanta, GA in collecting these data. Thanks to Duane Rumbaugh, Diana Reiss, Patricia Greenfield, David Washburn, and Irene Pepperberg for comments on earlier drafts. This research was supported by NICHD Grant 06016 to E. Sue Savage-Rumbaugh and the Language Research Center, Atlanta, GA. The author declares that she has no professional or financial affiliations that may be perceived to have biased this study. All applicable guidelines regarding animal care and use were followed in this experiment.

References

- Anwar MS (1979) Remarks on a collection of speech errors. *Int J Psycholinguist* 6:59–72
- Berg T (1996) The modality-specificity of linguistic representations: evidence from slips of the tongue and the pen. *J Pragmat* 27:671–697
- Berg T (2005) A structural account of phonological paraphasias. *Brain Lang* 94:104–129

- Bloom P (2000) How children learn the meanings of words. MIT, Cambridge
- Bovet D, Vauclair J (1998) Functional categorization of objects and of their pictures in baboons (*Papio anubis*). *Learn Motiv* 29:309–322
- Bovet D, Vauclair J, Blaye A (2005) Categorization and abstraction abilities in 3-year-old children: a comparison with monkey data. *Anim Cogn* 8:53–59
- Brakke KE, Savage-Rumbaugh ES (1995) The development of language skills in bonobo and chimpanzee-1. *Comprehension. Lang Commun* 15:121–148
- Brakke KE, Savage-Rumbaugh ES (1996) The development of language skills in *Pan* - II. Production. *Lang Commun* 16:361–380
- Brown DA, Boysen ST (2000) Spontaneous discrimination of natural stimuli by chimpanzees (*Pan troglodytes*). *J Comp Psychol* 114:392–400
- Calvin WH, Bickerton D (2000) *Lingua ex machina: reconciling Darwin and Chomsky with the human brain*. MIT, Cambridge
- Chomsky C (1969) The acquisition of syntax in children from 5 to 10. MIT, Cambridge
- Crowson K (1994) Errors made by deaf children acquiring sign language. *Early Child Dev Care* 99:63–78
- Cutler A (1982) *Speech errors: a classified bibliography*. Indiana University Linguistics Club, Bloomington
- Fromkin VA (1971) The non-anomalous nature of anomalous utterances. *Language* 47:27–52
- Fromkin VA (1973) Introduction. In: Fromkin VA (ed) *Speech errors as linguistic evidence*. The Hague, Mouton, pp 1–45
- Fromkin VA (1980) Introduction. In: Fromkin VA (ed) *Errors in linguistic performance: slips of the tongue, ear, pen and hand*. Academic, New York, pp 1–12
- Fromkin VA (1988) Grammatical aspects of speech errors. In: Frederick JN (ed) *Linguistic theory: extensions and implications*. Linguistics: The Cambridge survey, Vol 2. Cambridge University Press, Cambridge, pp 117–138
- Gardner RA, Gardner B T (1969) Teaching sign language to a chimpanzee. *Science* 165:664–672
- Gardner RA, Gardner B T (1984) A vocabulary test for chimpanzees (*Pan troglodytes*). *J Comp Psychol* 98:381–404
- Gelman SA, Wilcox S A, Clark E V (1989) Conceptual and lexical hierarchies in young children. *Cogn Dev* 4:309–326
- Greenfield PM (1980) Toward an operational and logical analysis of intentionality: the use of discourse in early child language. In: Olson D (ed) *The social foundations of language and thought: essays in honor of J. S. Bruner*. W. W. Norton & Co., New York
- Greenfield PM, Savage-Rumbaugh ES (1990) Grammatical combination in *Pan paniscus*: processes of learning and invention in the evolution and development of language. In: Parker ST, Gibson KR (ed) "Language" and intelligence in monkeys and apes: comparative developmental perspectives. Cambridge University Press, New York, pp 540–578
- Greenfield PM, Savage-Rumbaugh ES (1991) Imitation, grammatical development, and the invention of protogrammar by an ape. In: Krasnegor NA, Rumbaugh DM, Schiefelbusch RL, Studdert-Kennedy M (ed) *Biological and behavioral determinants of language development*. Lawrence Erlbaum, Hillsdale, pp 235–258
- Greenfield PM, Lyn H (2007) Symbol combination in *Pan*: language, action, and culture. In: Washburn DA (ed) *Primate perspectives on behavior and cognition*. American Psychological Association, Washington DC, pp 255–267
- Lazareva OF, Freiburger KL, Wasserman EA (2004) Pigeons concurrently categorize photographs at both basic and superordinate levels. *Psychon Bull Rev* 11:1111–1117
- Levy Y, Schlesinger IM (1988) The child's early categories: approaches to language acquisition theory. In: Levy Y, Schlesinger IM, Braine MDS (ed) *Categories and processes in language acquisition*. Lawrence Erlbaum, Hillsdale, pp 261–276
- Lyn H, Savage-Rumbaugh ES (2000) Observational word learning by two bonobos: ostensive and non-ostensive contexts. *Lang Commun* 20:255–273
- MacKay DG (1980) Speech errors: retrospect and prospect. In: Fromkin VA (ed) *Errors in linguistic performance: slips of the tongue, ear, pen and hand*. Academic, New York, pp 319–332
- Madole KL, Oakes LM (1999) Making sense of infant categorization: stable processes and changing representations. *Dev Rev* 19:263–296
- Mandler JM, Bauer PJ (1988) The cradle of categorization: is the basic level basic? *Cogn Dev* 3:247–264
- Mandler J M, McDonough L (1993) Concept formation in infancy. *Cogn Dev* 8:291–318
- Mareschal D, Quinn PC (2001) Categorization in infancy. *Trends Cogn Sci* 5:443–450
- Markman EM (1989) *Categorization and naming in children: problems of induction*. MIT, Cambridge
- Mayberry RI (1995) Mental phonology and language comprehension, or what does that sign mistake mean? In: Karen Emmorey JSR (ed) *Language, gesture, and space*. Lawrence Erlbaum, Hillsdale, pp 355–370
- Mervis CB (1987) Child-basic object categories and early lexical development. In: Ulric N (ed) *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Emory symposia in cognition, 1. Cambridge University Press, New York, pp 201–233
- Moran LJ (1973) Comparative growth of Japanese and North American cognitive dictionaries. *Child Dev* 44:862–865
- Murai C, Kosugi D, Tomonaga M, Tanaka M, Matsuzawa T, Itakura S (2005) Can chimpanzee infants (*Pan troglodytes*) form categorical representations in the same manner as human infants (*Homo sapiens*)? *Dev Sci* 8:240–254
- Nelson K (1974) Concept, word, and sentence: interrelations in acquisition and development. *Psychol Rev* 81:267–285
- Nelson K (1977) The syntagmatic-paradigmatic shift revisited: a review of research and theory. *Psychol Bull* 84:93–116
- Pepperberg IM (1987) Acquisition of the same/different concept by an African Grey parrot (*Psittacus erithacus*): learning with respect to categories of color, shape, and material. *Anim Learn Behav* 15:423–432
- Pepperberg IM (1992) Proficient performance of a conjunctive, recursive task by an African gray parrot (*Psittacus erithacus*). *J Comp Psychol* 106:295–305
- Perneger TV (1998) What is wrong with Bonferroni adjustments. *Br Med J* 136:1236–1238
- Pinker S (1994) *The language instinct*. W. Morrow and Co., New York
- Quinn P C, Eimas P D, Tarr M J (2001) Perceptual categorization of cat and dog silhouettes by 3- to 4-month-old infants. *J Exp Child Psychol* 79:78–94
- Roberts WA, Mazmanian DS (1988) Concept learning at different levels of abstraction by pigeons, monkeys, and people. *J Exp Psychol Anim Behav Process* 14:247–260
- Rosch E (1988) Principles of categorization. In: Collins AM, Smith EE (eds) *Readings in cognitive science: a perspective from psychology and artificial intelligence*. Morgan Kaufmann, San Mateo, pp 312–322
- Rosch EH (1973) On the internal structure of perceptual and semantic categories. In: Moore TE (ed) *Cognitive development and the acquisition of language*. Academic, New York, pp 111–144
- Rumbaugh DM (1977) *Language learning by a chimpanzee: the Lana project*. Academic, New York
- Savage-Rumbaugh ES, Rumbaugh DM, Smith ST, Lawson J (1980) Reference: the linguistic essential. *Science* 210:922–925
- Savage-Rumbaugh ES (1986) *Ape language: from conditioned response to symbol*. Columbia University Press, New York

- Savage-Rumbaugh ES, McDonald K, Sevcik RA, Hopkins WD, Rupert E (1986) Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *J Exp Psychol Gen* 115:211–235
- Savage-Rumbaugh ES, Murphy J, Sevcik RA, Brakke KE, Williams SL, Rumbaugh DM (1993) Language comprehension in ape and child. *Monogr Soc Res Child Dev* 58:v–221
- Tanaka M (2001) Discrimination and categorization of photographs of natural objects by chimpanzees (*Pan troglodytes*). *Anim Cogn* 4:201–211 10.1007/s100710100106
- Terrace HS (1979) *Nim. 1st Knopf*: distributed by Random House, New York
- Tomasello M (2003) *Constructing a language: a usage-based theory of language acquisition*. Harvard University Press, Cambridge
- Tomikawa SA, Dodd DH (1980) Early word meanings: perceptually or functionally based? *Child Dev* 51:1103–1109
- Vonk J (2002) Can orangutans (*Pongo abelii*) and gorillas (*Gorilla gorilla gorilla*) acquire concepts for social relationships? *Int J Comp Cogn* 15:257–277
- Vonk J, MacDonald SE (2002) Natural concept formation in a juvenile gorilla (*Gorilla gorilla gorilla*) at 3 levels of abstraction. *J Exp Anal Behav* 78:315–332
- Vonk J (2003) Gorilla (*Gorilla gorilla gorilla*) and orangutan (*Pongo abelii*) understanding of first- and second-order relations. *Anim Cogn* 6(2):77–86
- Vonk J, MacDonald SE (2004) Levels of abstraction in orangutan (*Pongo abelii*) categorization. *J Comp Psychol* 118:3–13
- Wallman J (1992) *Aping language*. Cambridge University Press, Cambridge
- Young ME, Wasserman EA (2001) Evidence for a conceptual account of same-different discrimination learning in the pigeon. *Psychon Bull Rev* 8:677–684