J. Child Lang. 17 (1990), 185–203. Printed in Great Britain

The comprehension of metaphor by preschool children*

BARBARA ZURER PEARSON
University of Miami

(*Received* 4 *January* 1988. *Revised* 15 *February* 1989)

ABSTRACT

Comprehension of metaphor in preschoolers was studied through an elicited repetition task. Subjects were 52 children aged 3;0 to 5;2. Repetition performance on metaphors was compared to repetitions of semantically well-formed literal sentences as well as semantically anomalous sentences, all matched for length, vocabulary and sentence structure. Accuracy on literal and metaphoric stimuli was comparable, and both were significantly better than performance on anomalous sentences. There were no effects for age or sex. It was shown that the metaphors were not semantically anomalous to the children and that they were processed on a par with literal language. The argument is advanced from a review of the literature that imitation implicates understanding of the material imitated. If metaphor is thus shown to emerge early in the child's linguistic repertory, figurative language, it may be argued, occupies a more central position in linguistic theory than it has been accorded.

INTRODUCTION

In the debate over the relation between literal and metaphoric language processing, child language may provide a key argument for one position over the other. On the one hand is the claim that literal language is basic and that metaphoric language represents an overlay, an added dimension or perhaps a systematic modification of the literal. In this view, we will expect the child to master metaphor well after the mastery of literal language. A contrasting position sees metaphor as a central, not marginal element, one that must be accounted for by any general theory of language processing. In this second scenario, metaphor will emerge early in the child's linguistic repertory.

[*] Address for correspondence: Barbara Pearson, Department of English, University of Miami, Box 248145, Coral Gables, FL 33124, USA. This research was supported by grants from the US Department of Education National Graduate Fellows Program and from the National Council of Teachers of English. I wish to thank Kim Oller and Virginia Gathercole for their valuable suggestions and encouragement.

In general, production studies of metaphoric language tend to support the early-metaphor view, while comprehension studies favour the claim for late acquisition. The current work presents experimental evidence for early metaphor comprehension, and thus strengthens the argument for metaphor as a primary language process.

The paradoxical situation in the child literature can be summarized as follows: children are quite universally recognized as producing utterances which adults interpret as figurative. Chukovsky's collection (1963), several articles in professional journals (Carlson & Anisfeld, 1969; Horgan, 1981), and at least one doctoral thesis (Harris, 1982) document for us the novel word uses of children that are so common to the point of being typical: *Oh look, he's barefoot all over. The chimney is a house-hat. The bad just crawls right out of me.* The young speaker shows no hesitation in transferring words from one domain to another, here seeing a house as wearing clothes, regarding a quality as animate, or using a restricted term, bareFOOT, in a more general sense.

Young children seem quite comfortable with novel lexical usage, with extending a word's meaning depending on what makes sense in the situation. Educators expect it of them: the Suzuki music method designed for preschoolers, for example, routinely personifies the violin the child holds, and the child is urged to treat the strings as hot or cold or sometimes fragile. Mother Goose describes a candle as wearing a petticoat, the month of March as having the attributes of two different animals, the dish and the spoon as a pair of fugitives. These are not far-out stories appealing only to the aesthetically sensitive child: these are the basic texts of childhood.

Children's use of metaphor-like expressions has been readily shown. In experimental situations, children have without effort located where a mountain's eyes would be, or its knees (Gentner, 1977); they have given metaphorical sentence completions ('he was as quiet as...') (Gardner, Kirchner, Winner & Perkins, 1975); and they have engaged in a renaming game at the experimenters' request, replicating on demand a skill they perform spontaneously in symbolic play (Hudson & Nelson, 1984; Winner, McCarthy & Gardner, 1980) – a block becomes a car and is called one, a hat an umbrella and so on. Their metaphors have been differentiated from misnomers and overextensions (Winner, 1979) and their classification behaviour has been probed to establish that this renaming behaviour does in fact involve for the child a transfer of a word from one domain to another (Vosniadou & Ortony, 1983; Mendelsohn, Robinson, Gardner & Winner, 1984). Even some awareness of the 'pretend' status of their renamings has been shown – from behavioural clues, a giggle or a pretend voice to a direct acknowledgement that 'we were just making believe' (Billow, 1981).

It has been much harder, however, for young subjects to demonstrate their comprehension of metaphors within the controlled situation of an experiment. The preliminary psycholinguistic investigation by Asch & Nerlove (1960) shows 3-, 5- and 7-year-old subjects rejecting the extension of physical terms to psychological attributes (a *hard rock* versus a *hard man*) and failing to consider a figurative interpretation for them. Billow (1975), Smith (1976), and Cometa & Eson (1978), all following a Piagetian tradition, asked their subjects to paraphrase or explicate various metaphors, never with very much success. Their collective conclusion has been that children cannot perform the logical operations necessary to first discover the analogical relationship between two terms of a metaphor and then transfer the meaning of one to the domain of the other until they are past the stage of concrete operations, at around eleven years (Billow, 1975). In the face of children's spontaneous use of metaphors (above), however, this conclusion is clearly unsatisfactory. Before concluding that children cannot perform the task, we must question the validity of the methods of demonstration. Paraphrase and explanation, in particular, add an extra step to the task. While they both rely on comprehension, they require the child to go beyond understanding in producing an adequate response.

What is needed is an experimental means of uncovering children's competence in the figurative domain which does not itself add substantially to the difficulty of the task, either by completely divorcing the expression from its supporting context (Siltanen, 1981; Vosniadou, Ortony, Reynolds & Wilson, 1984) or by adding a metalinguistic dimension to the response (Winner *et al.*, 1980).

By using more indirect measures, recent research has shown metaphoric understanding in children well below eleven. Winner *et al.* (1980) showed that 6-year-olds could uncover the analogical relationship between terms better when they were presented as riddles or quasi-analogies than as metaphors: 'The skywriting was a scar marking the sky' versus 'A scar marks someone's skin and — marks the sky'. Waggoner, Messe & Palermo (1985) embedded the metaphors within six- or seven-sentence stories and then compared recall of the metaphors with recall of their literal 'equivalents'. They found no significant difference for the figurative type. They also tested comprehension of the material following the metaphors in the story and found once again that the children did as well with metaphors as with the literal language. Kogan & Chadrow (1986), using their Metaphoric Triads Task, showed that children as young as seven both identified and explained metaphorical pairings among the stimuli.

Only one major study, however (Vosniadou *et al.*, 1984) has worked with children as young as three. In their elaborate paradigm, children enacted short stories with props provided by the experimenters, demonstrating their understanding, non-verbally, as the stories unfolded. Even the 3-year-olds showed some understanding of the metaphoric material, but at that age there was a highly significant difference between material with high or low predictability. The endings the children enacted for the less probable metaphors matched the children's performance in a 'no-ending' control,

where the experimenter just asked the child to show 'what he or she thought might happen', instead of providing the ending. The performance on even the highly probable metaphors was only slightly different from the 'no-ending' control, leading us to question the contribution of the linguistic input to the overall understanding.

Our goal, therefore, was to find a comprehension task (1) for which the successful response could not be gleaned from contextual factors alone; (2) where comprehension was a major component of the response and the added demands of the task were relatively easy; (3) that could be used with children younger than those for whom metaphorical understanding had thus far been shown. One method that meets all three of these requirements is elicited repetition. In particular, experience with this experimental task over the last two decades shows its intimate relation to comprehension.

Elicited repetition

Elicited repetition as a measure is sensitive to both variables of the input AND the subject's representation of the stimuli. The response mode, though, makes relatively little demand on the subject. Researchers who have used the technique observe that when the length of the stimulus string exceeds the child's memory, there are several factors which indicate that the response is 'filtered through' the linguistic system. When given unfamiliar stimuli, the child will often simplify them to something he or she does know. In other cases, where they do understand the stimulus, they tend to give an equivalent sentence (Slobin & Welsh, 1973). Smith (1973), like Menyuk (1963), reports a strong tendency for the children to normalize ungrammatical stimuli – i.e. repeat not what they actually heard, but what they thought they should have heard. Also errors – failures to give exact repetitions – were not randomly distributed nor even located principally in the middle of the string (where errors on nonsense materials are greatest), but rather they clustered at the more complex structures of the sentence.

The layman's impression that imitation of text is a mechanical procedure that can be accomplished without understanding of the material imitated is not supported by the facts. While short sequences may be 'parroted' back relatively adequately, Gallimore & Tharp (1981) have shown with a regression analysis that auditory memory *per se* makes only a small contribution to successful performance, greatly overshadowed in predictive ability by knowledge of the grammar used in the passages. The two dozen or so studies in the literature (see Pearson, 1987 for a review) which address the question establish imitation (both naturally occurring and elicited) as non-progressive; that is, subjects generally cannot imitate what they have not already given evidence of in their spontaneous speech. The developmental sequence supported by most studies is comprehension before production (Lackner, 1968; Bloom, Hood & Lightbown, 1974; Moerk, 1977), with imitation

falling sometimes before production and sometimes after (Slobin & Welsh, 1973; Kuczaj & Maratsos, 1975; Johnson, 1985). One counterexample to this sequence, Fraser, Bellugi & Brown (1963), is very well known. In this experiment, young subjects were shown to imitate sentences earlier than they could demonstrate their understanding of them. As Brown (1973) himself points out, though, the comprehension task was based on unnaturally difficult sentences with a minimum of redundancy, so that study is perhaps misleading with respect to this point.

Elicited repetition has been used much less in semantic studies than syntactic. The classic study on this is Miller & Isard (1963), which established for college students the relative ranking (in terms of numbers of correctly imitated strings) of meaningful grammatical sentences over semantically anomalous strings over ungrammatical strings. Marks & Miller's (1964) follow-up used four kinds of strings, well formed, semantically anomalous (words switched between sentences), anagrams (words switched within the well-formed sentences) and word lists (scrambled anomalous sentences). Their tally of percentage of words correct showed an equivalent decrement as a result of both semantic and syntactic disordering. Similarly, Eilers (1975) found the introduction of a nonsense word depressed repetition performance in general. As in the syntactic studies mentioned above, Love & Parker-Robinson (1972) found a tendency to 'normalize' in semantically vacuous 'sentences'; i.e. the children tended to substitute real English words for the nonsense syllables in their strings. So, in general, repetition experience with semantically distorted material shows a similar pattern of responses, attesting to an effect of the semantic distortion and the ACTIVE role of the subject in the task.

It is well established, then, and predictable that anomalous material will be repeated more poorly than semantically well-formed sentences. The question posed by this study was whether metaphoric sentences, in repetition, behave like the semantically well-formed, literal sentences or like the anomalous. If it is true that children do not understand metaphor, we would expect that the metaphoric sentences would be repeated significantly worse than the literal sentences, perhaps as poorly as the anomalous ones. If on the other hand, the metaphoric sentences are not semantically anomalous to our subjects, i.e. if they are understandable to them, the significant difference will be between metaphoric and anomalous, and the literal and metaphoric will not differ significantly. We make no claim about whether the children distinguish the metaphors from the literal expressions.

METHOD

Subjects

Subjects were 52 normal healthy children enrolled in three preschools in Greater Miami. There were 26 girls and 26 boys whose ages ranged from 3;0

to 5;2. They were divided for the analysis into groups of 26 under four and 26 over four, mean ages 3;7 and 4;6. All children were proficient speakers of English. (Four additional children completed the warm-up, but declined to go further.)

Materials

There were three tapes of prerecorded stimuli, form A, form B and form C, each consisting of 30 sentences graduated in length from 3 morphemes to 12 morphemes. The three sets were drawn from the same pool of sentences by randomly choosing one sentence from each of three semantic types (metaphoric, literal and anomalous) at each morpheme length. Form A presented sentences in the order metaphoric, literal, anomalous; form B, literal, anomalous, metaphoric; and form C, anomalous, metaphoric, literal. Starting with the 3-morpheme sentences, each form presented three sentences at each length, one of each semantic type. For the most part the forms had different sentences as well as the different order of types. However, since the random selection was done with replacement, there was some overlap. Each form had at least one sentence in common with another form, so there was a way to check whether differences between forms were the result of the different order or the different sentences. (See 'Interactions' in Discussion, below.)

The three types of stimulus sentence were constructed in the following way. For each of 51 metaphors, culled from children's books and other experiments on figurative language, a literal sentence and an anomalous sentence were created, matched for length in morphemes, using Roger Brown's rough guidelines (1973: 54) and, as far as possible, using the same sentence structure and vocabulary. All content words and prepositions in the metaphors were re-used in other sentences of approximately the same length. (Only in the literal condition were there any words that were not used elsewhere. Any such unmatched word was checked in the *American Heritage Word Frequency Book* (Carroll, Davies & Richman, 1971) to be in the 5,000 most frequent words of English. No words in the metaphor condition or the anomalous condition (and some of them are strange, e.g. *bullet*, *milkweed*, *fog*) appear in only one condition, and most are re-used in several sentences.) Whenever possible the animacy of the subject was maintained in all conditions, according to the caution presented in Reyna (1985); i.e. instead of changing the subject to match the verb, the subject was given a new, appropriate verb.

Anomalous sentences were constructed using the same sentence structure (typically subject–verb–object–prepositional phrase) in the correct order and respecting agreements, etc., but they combined words that created impossible or hard-to-interpret situations, for example, *someone climbed a night ship in the refrigerator*; or *the rain ate with three voices*.

After the matching, the sentences were incorporated in a questionnaire, scrambled to obscure the relationships between them and printed with a 5-point scale to the side for judgements about whether they 'made sense literally' (1), 'made sense metaphorically' (3), or 'didn't make (much) sense' (5), with the intermediate numbers for in-between judgements.

Twenty subjects, aged 15–79, acquaintances and colleagues of the author, all native speakers of English (although some were also bilingual), filled out the questionnaire independently. Their responses were tallied so that each answer that differed from our expectation was recorded. For a sentence to count as a particular type, there had to be 75% agreement, either to the original expectation or to each other. Intermediate numbers were counted as either of the adjacent numbers, but large numbers of 2s and 4s for any one sentence were taken into consideration and those sentences were not used.

The survey eliminated seven metaphoric, one literal and eight anomalous sentences. Unlike the longer anomalous sentences, our respondents found metaphorical possibilities in the anomalous candidates of five morphemes or less. (We went ahead and used the shorter sentences anyway as part of the warm-up, but this inability to make the distinction between the two types of sentence caused us to eliminate the 2-year-olds we had originally included in the design. We did not feel we could make any inferences based on data from only the shorter sentences.) This left a pool of 137 sentences from which to construct the stimulus sets. The sentences from form A appear in the Appendix.

Procedures

Each child was asked to repeat 30 sentences (ten each of the three semantic types, metaphoric, literal or anomalous) spoken from one of the three tapes (forms A, B or C). Subjects were assigned to conditions (choice of form) randomly.

The children were tested individually by the author in a small room at their preschool. After a brief conversation to familiarize the child with the tape-recorder, the directions, and the rewards and to secure his or her cooperation, the child repeated first digits and then nonsense syllables after the researcher (and received the first reward, a sticker). This time was used also to listen for any language problems that might invalidate the testing. Then the child listened to a set of sentences from one of the three stimulus tapes. The tape was stopped after sentences 9, 21 and 30 to give the child a reward. The whole interview was tape-recorded (with a dual recorder) for later transcription.

During the session, sentences were presented just once, unless the researcher judged that there was a clear external reason to prevent an adequate hearing of the sentences. If the children hesitated, they would be

prompted with 'Tell me the words you heard' or 'Tell me the words you remember'. If the children gave only the end of the sentence, only what they produced on the first attempt would be counted, but they would be encouraged to 'Say all the words' or say 'Is that how it started?' so that they would not get a mind-set to give just the ends. Comments by the children about the sentences were not discouraged, but care was taken in the practice trials to make sure that repetitions were given first. Verbal reinforcement was given after all responses, regardless of accuracy, with the longer break periods noted above. When a child completely missed all three sentence types at one length, the session ended. Fifteen of the 52 children stopped before 30 and four heard 33 sentences, but everyone attempted equal numbers of all three kinds of sentence.

The tapes were transcribed by the author soon after the sessions. At the conclusion of all the taping, a count of the numbers of errors in each sentence was taken. Half the data were transcribed and evaluated a second time by a research assistant who was not informed as to the purpose of the experiment, and the two data sets were checked for inter-rater reliability.

Scoring

All morphemes in the children's responses that deviated from perfect repetition were considered. Of these, a few deviations were considered equivalent to the original expressions and not counted as errors, namely: correct contractions; minor word-order changes that did not affect meaning, such as particle movement; self-corrected errors; the use of an alternate form (like 'dived' for 'dove') (cf. Kuczaj & Maratsos, 1975; Miller, 1973).

We made a further subdivision of errors into major and minor, where the minor errors represented articles, bound morphemes and other substitutions or omissions that are more likely characteristic of the language of 3- and 4-year-olds and therefore probably not affected by the conditions of our experiment (Brown, 1973; Scholes, 1969; see Pearson, 1987 for a review). These represented an average of about 4·5 errors per child, regardless of sentence type. The counting of the minor errors, however, was problematic in cases where whole passages of the sentences were omitted. Since the pattern of results for the major errors was the same as for all errors together, the analysis of the total scores is what we have reported here.

Inter-rater reliability

Kronbach's alpha was computed to compare the results of the two raters. The correlation for the metaphor scores was 0·89, the literal scores 0·90, and the anomalous scores 0·90, ($p < 0{\cdot}001$ for all values).

To summarize the differences in the scores between the raters, there was

a tendency for the second rater to find more total errors or higher error means in general. This can perhaps be quite naturally explained by her lesser familiarity with the children and with the materials. The MANOVAS were done separately using the second rater's data set, and the identical effects and interactions were found (although the mean square values were of course slightly different). Discussion by the raters about several protocols showed that some of the differences in the scoring could have been minimized by conferencing, but others represented different perceptions of what the child had said. Since both sets of interpretations yielded similar results, no steps were taken to improve the correlations. Scores reported are for the author's ratings.

Design

There was one within-subjects variable, semantic type (of the sentences) with three levels: literal, metaphoric and anomalous. There were three between-subjects variables: age (three or four years), sex and form (A, B or C). The dependent measure was the number of errors made in the repetitions.

RESULTS

An analysis of variance was done. *Post-hoc* pairwise contrasts were performed for the major effect of semantic type. An interaction between semantic type and form that emerged from the total design was followed up with an analysis of simple effects for the form variable, and the error means for the sentences that appeared on more than one form were analysed separately. A hand analysis of response patterns was compiled, and a correlation computed for the difference between the literal and anomalous scores with both the digit span and age in months. The error means are presented in Table 1.

TABLE 1. *Error means by semantic type*

Type	Metaphoric		Literal		Anomalous	
Error means	16·25	7·4	15·92	8·3	27·77	10·9
	Mean	S.D.	Mean	S.D.	Mean	S.D.

For the error scores, the within-subjects effect for semantic type was significant ($F = 64{\cdot}52$, $p < 0{\cdot}001$). The difference between the metaphor score and the literal score was not significant ($F = 0{\cdot}24$, $p > 0{\cdot}5$). But the difference between the metaphor and anomalous scores and between literal and anomalous were both significant: (M/A, $F = 74{\cdot}82$, $p < 0{\cdot}001$; L/A, $F = 72{\cdot}52$, $p < 0{\cdot}001$).

In the sub-analysis of error scores into major and minor errors (in 'Scoring' above), the mean number of major, or meaning-distorting errors for the literal and metaphoric sentences was not statistically different. They were 11·6 for the metaphors and 10·1 for the literal ($F = 3{\cdot}02$, $p > 0{\cdot}05$).

For the full design (semantic type by age by sex by form), only the within-subjects effect (semantic type) was significant. No between-subjects effect reached significance.

Interactions

In tests involving the within-subjects means, there was one significant interaction: form by semantic type ($F = 4{\cdot}73$, $p < 0{\cdot}005$). This interaction is shown in Fig. 1.

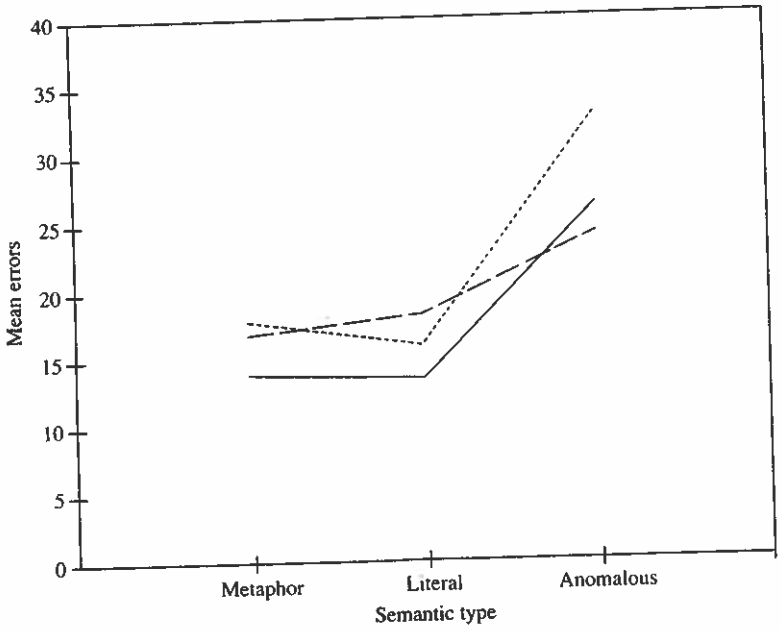

Fig. 1. Interaction: form by semantic type. —, Form A; ——, form B; ----, form C.

The simple effects were tested at each level of the form variable, and *t* tests were performed to find which differences were significant. These were all significant (A: $F = 20{\cdot}86$, $p < 0{\cdot}001$; B: $F = 10{\cdot}28$, $p < 0{\cdot}001$; C: $F = 48{\cdot}33$, $p < 0{\cdot}001$) and the *t* tests showed the difference between the metaphoric and literal sentences to be not significant, whereas the differences between each of those and the anomalous sentences were. For form B (where, as the chart in Fig. 1 shows, the effect was the least pronounced and hence, more likely

not to be significant) the *t*-values were as follows: M/L: $t = 0{\cdot}94$, $p = 0{\cdot}36$; M/A: $t = -3{\cdot}79$, $p = 0{\cdot}002$; L/A: $t = 3{\cdot}41$, $p = 0{\cdot}004$.

There are two possible sources of this interaction. The three forms were generated by a random drawing from the pool of matched sentences at the called-for lengths. There was some overlap, but for the most part the sentences were distinct. A second difference between the forms was the order of presentation – which type of sentence was first at each length. This possible confounding was deemed acceptable in order to broaden by almost three times the number of sentences that the study was based upon, and also to avoid further complicating the design for what was in essence a control variable.

As a check on whether the interaction was more likely a result of the different sentences or the order, three sentences that appeared in two versions each of the tapes were separately tabulated. One literal sentence that appeared on both forms B and C had more errors from the same number of children when it appeared on B than on C (45 to 37); the anomalous sentence on both A and C had more errors on C than on A; and the metaphor on C and A had more errors on C than on A – all in the direction of the overall form differences. This also suggests that the different response patterns were not so much a function of the different sentences as the presentation order.

The biggest effect by far was the poorer performance on the anomalous sentences in the anomalous-first condition. This can perhaps be explained by the mind-set this order may have established for the subjects. When the first sentence did not make sense, the children may have started out not treating the sentences to their best concentration. Performance on the metaphor, the next sentence, was also somewhat depressed, although the literal sentences were repeated best of any condition: perhaps the contrast was more striking in this direction. The best performance (fewest mean errors) on both metaphors and literal sentences was in the condition where they both preceded the anomalous (form A). This form may have established a task that was more reasonable to the children.

To summarize, the interaction is fairly minor compared to the robust main effect for semantic type that overrode potential age, sex or form effects. If anything, the control variable seemed to increase the magnitude of the central effect, but even without that enhancement the effect is very significant.

Response patterns

In addition to the statistical analyses, a hand check was done. In administering the protocols, we had noticed two distinct patterns of responses. Some children giggled or rolled their eyes at the sentences. As long as they gave the repetitions first, they were not discouraged from commenting between trials. A few subjects seemed unusually sensitive to the anomaly of

the sentences: one little girl declared *I can't say that* for each anomalous sentence; *That's silly*, said several others. One boy kept explaining why what the sentence said was impossible. By contrast, there were others who made no response to the sentences except to dutifully try to repeat, no matter how little they remembered or how little sense the fragments made to them. In looking back through all the scored protocols, we tried to determine how prevalent the different types were.

As predicted by the statistics, the majority of the subjects (29 out of 52) showed a clear pattern of approximately equal metaphor and literal errors, and a much larger error score for anomalous sentences. Eleven subjects had essentially equal error scores on all sentence types, either equally many or equally few errors. (Scores on the different sentence types were within six points and the greatest difference was not more than 15% of the total error score.) Of the other 12 subjects, who did not fall into either group, six had the fewest errors on the metaphors and six had their lowest scores on the literal sentences. (No subject did best on the anomalous sentences.)

Correlations

Also in keeping with the statistical result, no correlation was found with the different response patterns. In addition to the lack of main effect for sex, age

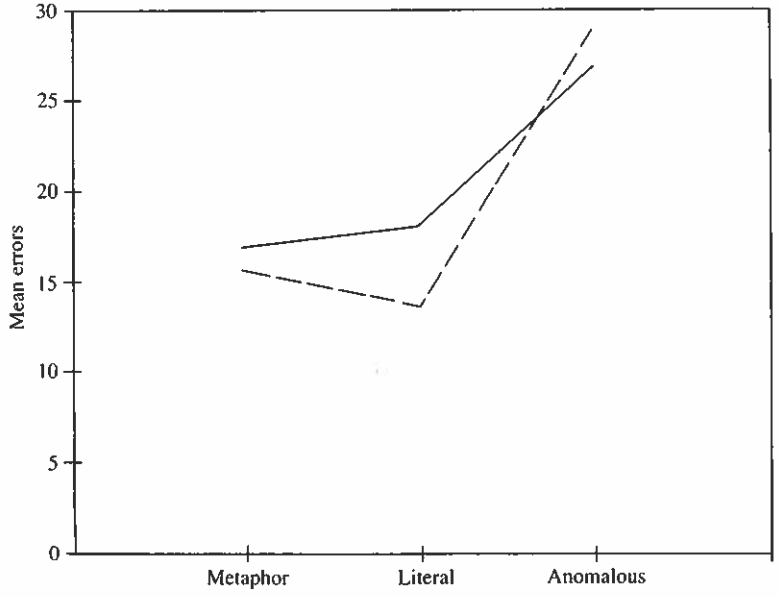

Fig. 2. Interaction: age by semantic type. —, 3 years; ---, 4 years.

and form, there was no relation of response pattern to the memory measure. Most of the children were able to repeat back four digits (30 out of 52); those who repeated five or six digits and those with fewer digits were fairly evenly distributed in either group. There was a slight tendency for 3-year-olds' scores to be relatively more even on the three types, whereas the 4-year-olds had a slightly lower mean for literal errors and slightly higher for anomalous (see Fig. 2), but a correlation between age in months and the anomalous–literal score difference was very low (r on the order of 0·14), and of course the interaction of age and semantic type did not reach significance.

All in all, there was a robust general effect in the direction of our original hypothesis that between the poles of literal and anomalous sentences, metaphor would align itself with the literal.

DISCUSSION

What does this result allow us to conclude? First, that children are not strangers to metaphor. They are here seen processing the figurative sentences on a par with the literal.

Further, we did not see a development with age in the capacity to repeat metaphors. There is so much individual variation in repetition ability in general that 4-year-olds were not consistently better than 3-year-olds. As for the contribution of memory, our finding is much the same as Gallimore & Tharp's (1981) regression analysis that memory *per se* makes only a small contribution to this performance. Four was the mode for the digit span at both ages, and the 3-year-olds had as many digit spans longer than four as the 4-year-olds. Four-year-olds did not appear to improve in their handling of metaphors over the 3-year-olds, although there was a slight suggestion that they may have been more sensitive to the anomaly in the third type of sentence when it appeared, perhaps a function of a growing metalinguistic awareness which is thought to be developing more rapidly at that age.

Relative performance

The strongest argument is made from the RELATIVE performance of the children with respect to the two opposing types of materials: the semantically well-formed literal sentences versus the semantically ill-formed sentences, judged by adults to 'not make sense'. According to the children's repetitions, the metaphors were in the category of 'making sense'.

Although many of the children volunteered that they were aware of the strangeness of the anomalous sentences, we have no evidence of their distinguishing the well-formedness of the metaphors and the literal sentences. That seems to us an interesting, but separate question, not tested directly in this paradigm. What was tested here is whether the children

distinguished the metaphors from the anomalous sentences, and clearly they did so. Since the sentences were constructed to differ only in semantic well-formedness, it is fair to say that the differences in the children's performance are in response to the degree of well-formedness. Such performance tells us, therefore, that the children treat the metaphors as well formed.

The ill-formedness of the anomalous sentences interfered with the children's ability to reproduce the sentences. Whether that was a fault of comprehension or memory was not formally tested. There were several indications, though, that the children did understand the sentences in general. For example, there were two imperatives among the four-morpheme sentences ('Turn off the radio/Turn off the fork'), which many of the children tried to comply with. We also observed chuckles, grimaces and even some exclamations in response to the sentences.

Also, as indicated above, we made an attempt to distinguish meaning-preserving errors from those that distorted meaning. The system we devised (described in more detail in Pearson, 1987) broke down because longer passages were totally omitted in the anomalous sentences, so that there was no way to classify the error. For the literal and metaphoric sentences, though, the pattern of no difference held up. The same percentage of errors seemed to us unrelated to meaning in the metaphors as in the literal sentences.

Next we looked through the data for instances where the children tried to 'normalize' the stimuli, perhaps making a metaphor or an anomalous sentence into a literal one. The most promising example of this was that five of 17 children substituted *frog* for *fog* in *the fog came in on little cat feet*; but *fog* was likewise mistaken five of 17 times, again three times with *frog*, in the literal foil, *The fog comes in after the rainstorm*. So, the difficulty seems more a function of the word than the semantic type.

We interpret the children's performance on the literal sentences as a kind of 'baseline' for their performance on this task in this setting. Thus, we have given evidence of children processing metaphors on a par with material they are thought to understand, simple literal sentences.

Relation of imitation to comprehension

Further corroboration of understanding can be gained from a consideration of the developmental sequence, where comprehension has been shown to precede imitation ability. This sequence has been established by direct comparison, i.e. by measuring comprehension and imitation with the same subjects and the same materials (Lackner, 1968). More often, though, imitated utterances have been compared on various measures – MLU, word order, presence of function words – to independent samples of spontaneous speech. Brown & Fraser (1963), Menyuk (1963), Prutting, Gallagher & Mulac (1976) and Eilers (1975) are among those who observed elicited

imitations, while Ervin (1964), Kamitomo & Ruder (1976), Kemp & Dale (1973), Snow (1981) and Bloom *et al.* (1974) observed the relationship with spontaneous imitations. While spontaneous imitation has shown some individual variation, the conclusion is almost unanimous that for most children imitation performance mirrors production.

From a couple of very interesting studies which integrate imitation into longitudinal designs focused on specific grammatical structures (Kuczaj *et al.*, 1975; Moerk, 1977; Johnson, 1985), we learn that the relationship between the two skills is more complex. Here a separate pattern emerged for imitation of grammatical and non-grammatical stimuli depending on the child's progress in the acquisition of the structure under study: the imitation was taken as illuminating an underlying competence which was not apparent from examination of spontaneous speech alone. Therefore, these studies do not weaken the argument for comprehension preceding imitation; rather, in so much as they too use imitation as evidence of understanding, they reinforce the general pattern.

Evidence for imitation behaviour being less advanced than production comes from two more naturalistic studies (Slobin & Welsh, 1973; Hood & Schieffelin, 1978) where the stimuli for imitation were the child's own utterance from conversations recorded earlier. The child's failure to repeat his or her own utterance can be traced perhaps to the added difficulty caused by isolating an utterance from the conversational supports of context and the motivation to communicate.

There have been some suggestions that repetition is progressive lexically (rather than syntactically). Moerk (1977: 69) found that words imitated spontaneously tended to be longer in syllables than his subjects' spontaneous usage, rarer and generally not in their lexicons, but the phenomenon was too limited for him to test it statistically. Bloom *et al.* (1974) and Scholes (1970) have also made allusions to a lexical motivation for imitation, but not in specific terms and not as a 'finding', so we are unable to identify from their speculations in what way such lexical progress could be measured. In order to take such a possibility into account, our study carefully balanced lexical items in all conditions.

While the evidence placing imitation after comprehension is not unequivocal, in the setting of this experiment, the difficulty of both tasks is increased by the researcher being a relative stranger to the child and by the lack of a supportive context of conversation or communicative intent. There seems no reason to think that one is facilitated over the other: difficulty in imitation will still implicate difficulty in understanding and, conversely, ability to imitate will depend crucially on ability to understand.

Relation to linguistic theory

Finally, we are not challenging Piaget's claims about the development of logical competence. It is no doubt true that to explain metaphors a child must have achieved the stage of 'formal operations' (Winner, 1979), but that does not entail that the child must have reached the same cognitive stage to use and understand them. Indeed, in the second year of life the prevalence of symbolic play assures us that conceptually children have the disposition to view one thing as another (Verbrugge, 1979) and in metaphor they have the linguistic means (Harris, 1982). Our study taps a skill that is metalinguistically less demanding than those used in previous studies and is thus, perhaps, closer to children's usage in naturalistic settings.

Figurative language in preschoolers is a topic which received relatively little attention in the 1960s and early 1970s. This neglect was due in part to an interpretation of Piaget (1962) which claimed that metaphor interpretation required children to have attained the stage of formal operations, and also perhaps because of Chomsky's insistence that figurative language was outside the scope of linguistic study (Chomsky, 1965; Katz & Fodor, 1963). Since the mid-1970s several major research groups have turned their attention to this topic, and its importance both for education and the theory of language is being recognized.

This experiment is simple and direct. The demonstration that metaphor is processed as well as equivalent literal sentences and better than anomalous strengthens the case for comprehension of figurative language by preschoolers. By lowering the age at which speakers are shown to be successfully processing metaphors, it moves figurative language toward its central place in language processing in general.

REFERENCES

Asch, S. & Nerlove, H. (1960). The development of double function terms in children. In B. Kaplan and S. Wapner (eds), *Perspectives in psychological theory*. New York: International Universities Press.

Billow, R. M. (1975). Cognitive developmental study of metaphor comprehension. *Developmental Psychology* **11**. 415–23.

—— (1981). Observing spontaneous metaphor in children. *Journal of Experimental Child Psychology* **31**. 430–45.

Bloom, L., Hood, L. & Lightbown, P. (1974). Imitation in language development; if, when and why. *Cognitive Psychology* **6**. 380–420.

Brown, R. (1973). *A first language: the early stages*. Cambridge, MA: Harvard University Press.

Brown, R. & Fraser, C. (1963). The acquisition of syntax. In C. Cofer & B. Musgrave (eds), *Verbal behavior and learning: problems and processes*. New York: McGraw-Hill.

Carlson, P. & Anisfeld, M. (1969). Some observations on the linguistic competence of a two-year-old child. *Child Development* **40**. 569–75.

Carroll, J. B., Davies, P. & Richman, B. (1971). *Word frequency book*. New York: American Heritage.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge MA: M.I.T. Press.

Chukovsky, K. (1963). *From two to five*. Berkeley CA: University of California Press.

Cometa, M. S. & Eson, M. E. (1978). Logical operations and metaphor interpretation: a Piagetian model. *Child Development* **49**. 649–59.

Eilers, R. (1975). Suprasegmental and grammatical control over telegraphic speech in young children. *Journal of Pyscholinguistic Research* **4**. 227–39.

Ervin, S. (1964). Imitation and structural change in children's language. In E. Lenneberg (ed.), *New directions in the study of language*. Cambridge MA: M.I.T. Press.

Fraser, C., Bellugi, U., & Brown, R. (1963). Control of grammar in imitation, comprehension, and production. *Journal of Verbal Learning and Verbal Behavior* **2**. 121–35.

Gallimore, R. & Tharp, R. (1981). The interpretation of elicited sentence imitation in a standardized context. *Language Learning* **31**. 369–92.

Gardner, H., Kircher, M., Winner, E. & Perkins, D. (1975). Children's metaphoric productions and preferences. *Child Language* **2**. 125–41.

Gentner, D. (1977). Children's performances on a spatial analogies task. *Child Development* **48**. 1034–9.

Harris, P. G. (1982). The development of the use and comprehension of metaphor by preschool children. Unpublished doctoral dissertation, Texas A&M.

Hood, L. & Schieffelin, B. (1978). Elicited imitation in two cultural contexts. *San Diego Laboratory of Comparative Human Cognition Quarterly Newsletter* **2**. 4–12.

Horgan, D. (1981). Learning to tell jokes: a case study of metalinguistic abilities. *Journal of Child Language* **8**. 217–24.

Hudson, J. & Nelson, K. (1984). Play with language: overextensions as analogies. *Journal of Child Language* **11**. 337–46.

Johnson, C. (1985). The emergence of present perfect verb forms: semantic influences on selective imitation. *Journal of Child Language* **12**. 325–52.

Kamitomo, G. H. & Ruder, K. F. (1976). Young children's spontaneous imitations of well-formed and anomalous commands. Unpublished master's thesis, University of Kansas.

Katz, J. & Fodor, J. (1963). The structure of a semantic theory. *Language* **39**. 170–210.

Kemp, J. & Dale, P. (1973). Spontaneous imitation and free speech: a developmental comparison. Paper presented at the Biennial Conference of the Society for Research in Child Development, Philadelphia.

Kogan, N. & Chadrow, M. (1986). Children's comprehension of metaphor in the pictorial and verbal modality. *International Journal of Behavioral Development* **9**. 285–95.

Kuczaj, S. & Maratsos, M. (1975). What children can say before they will. *Merrill Palmer Quarterly* **21**. 89–111.

Lackner, J. R. (1968). A developmental study of language behavior in retarded children. *Neuropsychologia* **6**. 301–20.

Love, J. M. & Parker-Robinson, C. (1972). Children's imitation of grammatical and ungrammatical sentences. *Child Development* **43**. 309–19.

Marks, L. E. & Miller, G. (1964). The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior* **3**. 1–5.

Mendelsohn, E., Robinson, S., Gardner, H. & Winner, E. (1984). Are preschoolers' renamings intentional category violations? *Developmental Psychology* **20**. 187–92.

Menyuk, P. (1963). A preliminary evaluation of grammatical capacity in children. *Journal of Verbal Learning and Verbal Behavior* **2**. 429–39.

Miller, G. & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior* **2**. 217–28.

Miller, J. F. (1973). Sentence imitation in preschool children. *Language and Speech* **16**. 1–14.

Moerk, E. L. (1977). Processes and products of imitation: additional evidence that imitation is progressive. *Journal of Psycholinguistic Research* **6**. 187–202.

Pearson, B. Z. (1987). Review of elicited repetition experiments. Unpublished manuscript, University of Miami.

Piaget, J. (1962). *Plays, dreams and imitation in childhood*. New York: Norton.

Prutting, C. A., Gallagher, T. M. & Mulac, A. (1976). The expressive portion of the NSST compared to a spontaneous language sample. *Journal of Speech and Hearing Disorders* **40**. 40–8.

Reyna, V. F. (1985). Figure and fantasy in children's language. In M. Pressley & C. Brainerd (eds), *Cognitive learning and memory in children*. New York: Springer-Verlag.
Scholes, R. J. (1969). The role of grammaticality in the imitation of word strings by children and adults. *Journal of Verbal Learning and Verbal Behavior* **8**. 225–8.
—— (1970). On functors and contentives in children's imitations of word strings. *Journal of Verbal Learning and Verbal Behavior* **9**. 167–70.
Siltanen, S. A. (1981). Apple noses and popsicle toeses: a developmental investigation of metaphorical comprehension. Unpublished doctoral dissertation, Ohio State University.
Slobin, D. I. & Welsh, C. A. (1973). Elicited imitation as a research tool in developmental linguistics. In C. A. Ferguson & D. I. Slobin (eds), *Studies of child language development*. New York: Holt, Rinehart & Winston.
Smith, C. (1973). An experimental approach to children's linguistic competence. In C. A. Ferguson & D. I. Slobin (eds), *Studies of child language development*. New York: Holt, Rinehart & Winston.
Smith, J. W. A. (1976). Children's comprehension of metaphor: a Piagetian interpretation. *Language and Speech* **19**. 236–43.
Snow, C. (1981). The uses of imitation. *Journal of Child Language* **8**. 205–12.
Verbrugge, R. (1979). The primacy of metaphor in development. *New Directions for Child Development* **6**. 77–84.
Vosniadou, S. & Ortony, A. (1983). The emergence of the literal–metaphorical–anomalous distinction in young children. *Child Development* **54**. 154–61.
Vosniadou, S., Ortony, A., Reynolds, R. E. & Wilson, P. (1984). Sources of difficulty in the young child's understanding of metaphorical language. *Child Development* **55**. 1588–1606.
Waggoner, J. E., Messe, M. J. & Palermo, D. S. (1985). Grasping the meaning of metaphor: story recall and comprehension. *Child Development* **56**. 1156–66.
Winner, E. (1979). New names for old things: the emergence of metaphoric language. *Journal of Child Language* **6**. 469–91.
Winner, E., McCarthy, M. & Gardner, H. (1980). The ontogenesis of metaphor. In R. P. Honeck & R. R. Hoffman (eds), *Cognition and figurative language*. Hillsdale NJ: Erlbaum.

APPENDIX

Sentences on Form A

3 m[orphemes]:	Metaphor	(1)	The sky fell.
	Literal	(2)	Johnny ran away.
	Anomalous	(3)	Water exploded.
4 m:	M	(4)	The radio growled.
	L	(5)	The child was thirsty.
	A	(6)	Turn off the fork.
5 m:	M	(7)	Hear the wind singing.
	L	(8)	The water is running.
	A	(9)	Sing the wind sailing.
6 m:	M	(10)	Dawn creeps across the sky.
	L	(11)	The people ate everything all up.
	A	(12)	Fairies flow with the cigarette.
7 m:	M	(13)	The rainbow is a bridge in heaven.
	L	(14)	The man is the teacher's husband.
	A	(15)	The cat ties a loud snail.
8 m:	M	(16)	The trees stretch out their arms.
	L	(17)	The sound was the cry of his brother.
	A	(18)	The rains stretch it with songs.
9 m:	M	(19)	The stars are the moon's children.
	L	(20)	The fog comes in after the rain storm.

	A	(21)	Newspapers are stars wearing the bath.
10 m:	M	(22)	The daisies stick their toes in the ground.
	L	(23)	Teardrops made lines rolling down her face.
	A	(24)	The rainbow is a broom hiding without our arm.
11 m:	M	(25)	I am the lawn's barber cutting its green hair.
	L	(26)	Puddles in the street are too cold for my feet.
	A	(27)	The parades were forests climbing in their toe.
12 m:	M	(28)	He took the road with a mountain tied to its end.
	L	(29)	The man blew three puffs of smoke from his cigarettes.
	A	(30)	The southwind took sticks and kittens with a thick mountain.