

Interpreting Metaphors and Idioms: Some Effects of Context on Comprehension

ANDREW ORTONY, DIANE L. SCHALLERT, RALPH E. REYNOLDS, AND STEPHEN J. ANTOS

University of Illinois at Urbana-Champaign

Two experiments are described in which reaction times for understanding target sentences or phrases in terms of a preceding context were measured. In Experiment 1, the target sentences followed either short or long contexts which induced either literal interpretations or metaphorical ones. Results indicated that only in the short context condition did subjects take significantly longer to understand metaphorical than literal targets. This interaction is explained in terms of the availability of appropriate schemata for interpreting the target. In Experiment 2, targets were phrases that could be given either an idiomatic or a literal interpretation. It was found that the comprehension of phrases receiving an idiomatic interpretation took no longer than the comprehension of those same phrases when given a literal interpretation, and there was some evidence that idiomatic interpretations were consistently faster. It is argued that both experiments can be accounted for in terms of contextually generated expectations. The processes required for the comprehension of figurative and literal uses of language seem to be essentially similar.

While philosophers have been speculating about the nature of metaphor at least since the time of Aristotle, psychologists have only recently begun seriously to investigate it. Most of their work has been concerned with developmental trends and suffers from a variety of conceptual and methodological problems (see Ortony, Reynolds, & Arter, 1978). Yet metaphor is an important problem in cognitive psychology. Explaining metaphor constitutes a challenging test for theories of similarity judgments (e.g., Tversky, 1977) and analogical reasoning (e.g., Sternberg, 1977) as

We are grateful to Mark Anderson, Ernie Goetz, and Zohara Shifrin for their help in the early stages of this research, and to Richard Vondruska for his help with programming. We also wish to thank Glenn Kleiman for helpful comments on earlier drafts of this paper. The work was supported in part by the National Institute of Education under Contract No. US-NIE-C-400-76-0116, and by a Spencer Fellowship awarded to the first author by the National Academy of Education. Send requests for reprints to Andrew Ortony, Center for the Study of Reading, University of Illinois, 174 CRC, 51 Gerty Drive, Champaign, Illinois, 61820.

well as for theories of language comprehension in general.

One approach to the analysis of the comprehension of metaphors is the Pragmatics approach based on recent work of Grice (1975), Searle (1969; in press), and others, who suggest that linguistic communication is governed by a tacit agreement to cooperate in the communicative act. According to Grice, apparent violations of this agreement serve communicative functions. When a hearer (or a reader) discovers an violation, he or she tries to reinterpret what was said so as to render the violation only apparent. Typical examples of such violations occur in indirect speech acts where one kind of illocutionary force (e.g., that of assertion, as in, *It's cold in here*) is used to convey another kind of illocutionary force (e.g., that of a request, perhaps to close a window or to turn up the thermostat). Searle (in press) suggests that metaphors also involve such violations. He argues that the chief difference between the

two is that when using an indirect speech act the speaker intends to communicate both the literal meaning of what is said *and* the conveyed meaning. By contrast, when using a metaphor the speaker's goal is to convey only the metaphorical meaning. In both metaphors and indirect speech acts, the hearer's task is to figure out what the speaker meant (utterance meaning) from what he said (sentence meaning), given that the context makes it clear that the two are different, and that in that context the sentence meaning is defective. This "figuring out" entails that the hearer must first process the literal meaning of the utterance to a sufficient extent to determine whether or not it is compatible with the context. If it is not, the hearer has to engage in further processing to determine the utterance meaning.

Searle's analysis of the comprehension process for metaphors, therefore, entails three stages. First, the literal meaning of the utterance is determined. Second, that meaning is checked against the context. Third, if there is a conflict between the literal meaning and the context, it is reinterpreted and a conveyed meaning is derived. Clark and Lucy (1975) tested three predictions of just this model to determine whether it would provide a reasonable account of the comprehension process involved in sentences expressing indirect speech acts. One prediction was that subjects would show evidence of having processed the sentences literally. A second prediction was that indirect requests would take longer to be understood than direct requests. The third prediction was that subjects would show evidence of having derived a reinterpreted meaning. The subjects' task involved verifying whether direct and indirect, positive and negative requests had been fulfilled. Response times were compared with predicted patterns for verification times of positive and negative statements. Clark and Lucy interpreted their results as strong direct support for the first and third predictions, and as reliable indirect evidence for the second.

Focusing on metaphors, Verbrugge and

McCarrell (1977) presented subjects with a series of metaphorical sentences such as *Billboards are warts on the landscape*. They reasoned that comprehending these sentences would involve determining the "ground," namely, the shared meaning components of the metaphorical "topic" (in this case, *billboards*) and the metaphorical "vehicle" (in this case *warts*). They further argued that if comprehension did indeed require the determination of the ground then the ground should be a very effective retrieval cue. They found this to be so. Thus, for instance, the ground in the example just given, *ugly protrusion on a surface*, was a much better retrieval cue than was the ground, *tell you where to find businesses in the area*, which in turn was more effective for a control metaphor, *billboards are the yellow pages of a highway*. These results suggest that the comprehension of metaphors requires subjects to make inferences about what the ground of a metaphor is—inferences that would not be necessary in the comprehension of literal statements such as *Billboards are ugly protrusions on the landscape*. A reasonable extrapolation from their results is that the additional inferences necessary for the comprehension of metaphors will manifest themselves in increases in processing time.

In both lines of research described above, an increase in processing time for metaphors seems to be indicated even though the theoretical origins are rather different. Clark and Lucy can be regarded as testing a general model for the comprehension of nonliteral uses of language, which they claim involves three stages. Verbrugge and McCarrell can be regarded as investigating what is involved in the third, reinterpretation, stage. It has been argued, however, that the comprehension of metaphors does not involve stages of comprehension. Verbrugge (1977), for example, suggests that such a model is unparsimonious. He proposes instead that all language, metaphorical or literal, is understood through elaboration processes that are constrained by

the context. A similar position is taken by Rumelhart (in press). Harris (1976), in attempting to repudiate the stage model, found that subjects took no longer to initiate a paraphrase of a metaphor than they did to initiate a paraphrase of its literal equivalent, although questions arise over the appropriateness of such a dependent measure.

Our proposal is that the stage model is not incorrect but that it represents a limited rather than a general account of the comprehension of figurative language. In general, we propose that a hearer or reader uses an already constructed representation of what has gone before (the context) as a conceptual framework for interpreting a target sentence, or any other linguistic unit (see, e.g., Bransford & Johnson, 1972; Schallert, 1976). In many cases the interpretation is quite unproblematic; the reader or listener can almost predict what will be conveyed and the target sentence is used, as it were, to confirm an already formed hypothesis about its meaning. In other cases, however, for one reason or another, this fairly automatic, predominantly top-down strategy is not possible. Such a situation may exist when, as in the Clark and Lucy and the Verbrugge and McCarrell experiments, target sentences are presented with minimal or no preceding context. Without contextual support to guide expectations, the inferential processes entailed by the stage model and by Verbrugge and McCarrell's analysis may have to be made quite deliberately. Thus, our view is that whether or not a target sentence requires a relatively large amount of processing time is a function of how easily it can be interpreted in the light of contextually determined expectations rather than a function of its non-literalness. It is incorrect to assume that non-literalness always impairs the ease of interpretation. Our proposal seems to have the following empirical consequences, which were tested in Experiment 1. First, given insufficient contextual support, targets requiring a metaphorical interpretation should take longer to be processed than targets requiring a literal inter-

pretation. Second, given sufficient contextual support, they should not.

In the first experiment a number of "vignettes" were constructed. Each vignette comprised two parts, an antecedent context and a sentence-length target. For each item there were two kinds of contexts, one that was intended to induce a literal interpretation of the target (the literal inducing context) and one that was intended to induce a metaphorical interpretation of it (the metaphorical inducing context). The dependent measure was the time taken for a subject to understand the target. The procedure was, therefore, similar to that of Haviland and Clark (1974), not only with respect to the dependent measure but also in that antecedent context was manipulated to alter the interpretation of a target. Note that in contrast to the metaphors of Verbrugge and McCarrell which represented semantically anomalous uses of words, our metaphors were contextually anomalous whole sentences. An example of an item is given below:

Literal inducing context

Approaching the enemy infantry, the men were worried about touching off landmines. They were very anxious that their presence would be detected prematurely. These fears were compounded by the knowledge that they might be isolated from their reinforcements. The outlook was grim.

Metaphorical inducing context

The children continued to annoy their babysitter. She told the little boys she would not tolerate any more bad behavior. Climbing all over the furniture was not allowed. She threatened to spank them if they continued to stomp, run, and scream around the room. The children knew that her spankings hurt.

Target

Regardless of the danger, the troops, marched on.

Short contexts were produced from the long ones by using only the first sentence or part thereof. Thus, in the short context condition the target *Regardless of the danger, the troops*

marched on was preceded by either *Approaching the enemy infantry* or *The children continued to annoy their babysitter*.

To summarize our position, the general account of comprehension that we have offered suggests that whether one target will be comprehended more quickly than another depends not so much on whether it has to be interpreted literally but upon the extent to which it can be readily interpreted in terms of a reader's or listener's emerging representation of what has gone before. This leads to the following two hypotheses:

(1) *In cases where there is little contextual support*, sentences given a metaphorical interpretation (hereafter called "metaphors") require more time for their comprehension than sentences given a literal interpretation (hereafter called "literals"), i.e., the stage model provides an adequate account.

(2) *In cases where there is abundant contextual support*, metaphors are processed as quickly and easily as literals, i.e., the stage model does not provide an adequate account.

EXPERIMENT 1

Method

Design. The main part of the experiment involved measuring the amount of time it took for a subject to indicate that he had understood a particular sentence under different context conditions. Each person was assigned randomly to list (List A or List B) and order (Order 1 or Order 2) conditions. Context length (short or long), a between-subject factor, was randomly assigned to small groups of subjects being tested together. Context type (metaphorical or literal) was a within-subject factor.

Subjects. Forty students, 4 males and 36 females, were recruited from an undergraduate general educational psychology class and paid \$2.00 for their participation.

Materials. Each item consisted of a target sentence which could be interpreted either literally or metaphorically, depending upon the

context which preceded it. A particular target could appear after either a short or a long version of either a metaphorical-inducing or a literal-inducing context. The items were written in keeping with the following guidelines: (a) contexts should induce either clearly metaphorical or clearly literal interpretations of the targets; (b) the contexts themselves should be written using only literal language; (c) the target should not merely repeat or translate one of the context sentences but should be a continuation or summary sentence; and (d) the degree to which the target follows from the context should be as equal in the metaphor and literal versions as possible. How well the items adhered to these guidelines was tested by presenting 23 items to 140 undergraduate students enrolled in a Philosophy of Education course. For each target, half of the students read the metaphorical context and half read the literal context. Half of the subjects received short contexts and the rest long contexts. They were asked to indicate whether a particular target sentence should be interpreted literally, metaphorically, or whether it was uninterpretable given the preceding context. In addition, they were asked to indicate the degree to which a target was related to one of its contexts on a 7-point scale. Based on these ratings, the 16 items which were most consistently categorized as literal or metaphorical in their appropriate context conditions and which showed the highest degree of relatedness to the context were selected as experimental items. Twenty other items were then constructed and, together with the seven rejected items from the norming study, appeared as practice trials before the experimental items.

The 16 experimental target sentences ranged from 4 to 10 words in length with a mean of 7 words. Short contexts ranged from 3 to 11 words with a mean of 6, and long contexts ranged from 33 to 60 words with a mean of 45. For each item the average difference in length between metaphorical and literal contexts was 4 words for the long versions and 2 words for the short versions. Two lists were

constructed so that in each list, 8 of the targets appeared after metaphorical contexts and 8 appeared after literal contexts. Targets requiring a metaphorical interpretation in List A were interpretable literally in List B and vice versa. In addition, there was a short and a long context version for each list and two random orders of the items for each list.

Procedure. All subjects received the same instructions. They were told by an experimenter that they were participating in a study dealing with language comprehension. The general form of the items, a context section followed by a target sentence, was described although no mention was made of the metaphorical nature of some of the items. The general procedure for interacting with the computer terminal was then described. The experimenter explained that a ready signal would be on the screen when they sat down at the terminal. As soon as they were ready to begin, they should press the space bar on the keyboard in front of the screen, causing the first part of an item to appear. They were told that when they had read and understood this part, they should press the space bar again, causing the second part to appear. Then, as soon as they had read and understood this part, they should press the bar. After a 5-second interval, the ready signal would appear for the next item. The experimenter reminded the subjects that the target parts should be understood in terms of the context sections. He discouraged them from spending an inordinate amount of time on each section and from pressing the bar before they had understood what they had read.

Subjects were then escorted to individual sound-proofed cubicles, each equipped with a computer terminal consisting of a small CRT screen and a typewriter keyboard. The computer system, capable of handling several subjects simultaneously, regulated the presentation of all items and recorded the elapsed time in milliseconds between bar presses. All subjects received 27 practice items and then the 16 experimental items without interruption.

Upon completing the main part of the experiment, the subjects were taken to a small classroom where they were given a posttest. The test consisted of the 16 target sentences along with instructions to write a brief summary of how each of these had been interpreted in the first phase of the experiment. Subjects were discouraged from guessing or making up an interpretation on the spot. Note that 24 of the subjects participated in the main phase of Experiment 2 before going on to the posttests for both experiments.

Results and Discussion

Preliminary analyses were run which indicated that, within context length conditions, subjects did not differ in the amount of time they spent reading the metaphorical and literal inducing context sections, $F < 1$, in all cases. Also, order of items was not a significant factor nor did it interact with other factors in analyses of reaction times for the targets. Therefore, order was dropped as a factor in further analyses.

The amount of time it took for a subject to indicate that the predicted interpretation had been made and understood was the measure of prime interest. The data from three subjects were discarded because they did not have time to complete the posttest. Also, four data points which fell three standard deviations above individual subjects' mean reaction time on all targets were removed as well as four data points which were below 700 milliseconds. These outliers represented cases in which experimental instructions could not have been followed; a subject's attention may have wandered from the task, or he or she may have reacted accidentally before the target could possibly have been read and understood. In addition, since the reaction time measure did not allow a direct check of subjects' interpretations and since our predictions were based on clear cases of metaphorical and literal processing, answers on the posttest were examined and data points corresponding to targets which were clearly misinterpreted were

removed. Such cases, where subjects reported that they had interpreted literally targets that were meant to be metaphorical, or vice versa, amounted to a total of 22 cases (4%). In comparison, 65% of the posttest answers indicated correct interpretations, 28% fell in the "forgot" category, and 3% could not be categorized unambiguously. The scores trimmed from the data (a total of 30 data points or 5% of the total) were evenly distributed across all levels of the critical variables.

Analyses of variance were performed with context length (short vs long) and list (A vs B) as between-subject factors, and with context type (metaphor vs. literal) as a within-subject factor. Figure 1 represents the mean reaction times for interpreting targets in the various conditions. Results indicated a significant effect for context length, $F_1(1, 33) = 14.54$, $p < .001$, $F_2(1, 15) = 133.18$, $p < .001$, $minF'(1, 45) = 11.71$, $p < .01$, thus demonstrating that targets following long contexts were processed more quickly than targets following short contexts. The context type effect was also significant, $F_1(1, 33) = 20.92$, $p < .001$, $F_2(1, 15) = 26.62$, $p < .001$, $minF'(1, 40) = 13.11$, $p < .001$, with metaphors taking longer to be processed than literals. Finally,

the interaction of these two variables was significant, $F_1(1, 33) = 5.42$, $p < .05$, $F_2(1, 15) = 9.54$, $p < .01$. Although the $minF'$ for the interaction was only marginally significant, $minF'(1, 47) = 3.46$, $p < .075$, $maxF'$ did reach conventional levels of significance, $maxF'(1, 47) = 4.94$, $p < .05$.

Simple effects analyses of the interaction indicated no significant difference between metaphorical and literal targets in the long context condition, $F_1(1, 66) = 2.73$, $p > .10$, but a significant difference in the short context condition, with metaphorical targets being processed more slowly than literal targets, $F_1(1, 66) = 22.73$, $p < .001$, $F_2(1, 30) = 39.92$, $p < .001$, and $minF'(1, 95) = 14.48$, $p < .001$.

These results indicate that while subjects took longer to interpret targets in metaphorical than in literal contexts in the short context condition, there was no significant difference between metaphors and literals in the long context condition. Thus, the process of first interpreting a sentence literally, then determining that such an interpretation does not fit the context, and finally computing the intended figurative meaning does not seem to always underlie the interpretation of figurative language.

Surely the stage model does not apply to cases in which a conventionalized meaning of a nonliteral expression, such as an idiom, is highly determined by the context. For example, if the context sets up the expectation that something is rather irritating and annoying, the familiar expression, *a pain in the neck* (or other expressions comparably colloquial but a little less tasteful), would permit the immediate satisfaction of the contextually generated expectations. The use of such expressions is highly conventionalized and, indeed, is very comparable to the use of certain indirect speech act forms for which we would make similar predictions (e.g., *can you do x*, meaning, *please do x*; or, *do you have a/the x?*, meaning, *please give/pass/lend a/the x*). So, the account being proposed not only predicts

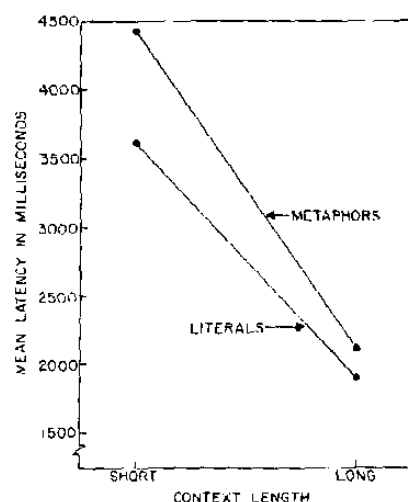


FIG. 1. Mean reaction times for understanding targets.

that under appropriate circumstances, as Clark and Lucy (1975) point out (p. 69), the stage model would not be supported, but it also predicts that, at least in some cases, non-literal uses of language might be processed faster than literal uses. Experiment 2 was designed to test this hypothesis.

EXPERIMENT 2

Method

Design. Again the measure of prime interest was the time it took subjects to indicate that they had understood a target phrase under different context conditions. Subjects were assigned randomly to three lists and to two different random orderings of each list. Within each list, there were three types of items: some for which the target was to be interpreted idiomatically (idioms); others for which the target was an idiomatic phrase but which should be interpreted literally given the context (literals); and finally, items which did not include any idiomatic expression but were literal paraphrases of the idioms. These control items were meant to be interpreted literally. A final factor, identified as sequence, related to whether subjects participated in Experiment 2 only (idioms-first condition), or whether they completed the main (reaction time) phase of Experiment 1 before seeing the main phase of Experiment 2 and then going on to complete the posttest for both experiments (idioms-second condition). The rationale for this factor was only that of increasing the efficiency of subject time. It was hoped that there would be no difference between the two conditions and that the power of the test for a context type effect would be increased because of the greater number of subjects.

Subjects. Students in an undergraduate educational psychology class were recruited and paid \$2.00 for participation. There were 48 subjects in this experiment, 40 females and 8 males.

Materials. Items were constructed in the same general style as items in Experiment 1, a

context section being immediately followed by a target. There were three versions of each item. In the idiom version, an idiomatic expression such as *let the cat out of the bag* appeared after a context which induced an idiomatic interpretation of the phrase. In the literal version, the same expression appeared following a context which encouraged a literal interpretation of the target. In the control version, the same context which appeared in the idiom version was followed by a target which was a literal paraphrase of the idiomatic expression. The following is an example of the three versions of one item:

Idiom version

Context: Dean spoiled the surprise that Joan had been planning for their mother's birthday party. When he realized what he'd done, he apologized for having

Target: let the cat out of the bag.

Literal version

Context: Walking back from the store, Anne found a kitten which she put in with her groceries. She got home and her puppy went wild when she

Target: let the cat out of the bag.

Control version

Context: Dean spoiled the surprise that Joan had been planning for their mother's birthday party. When he realized what he'd done, he apologized for having

Target: revealed the secret.

Items were written in keeping with the guidelines outlined in Experiment 1. However, in this experiment, two additional constraints were imposed. The first constraint was that each target should be comprised of the idiomatic expression or its literal translation alone. Second, the literal translation appearing as target in the control version was not to have more words than the idiomatic expression. Actual experimental items were selected following a norming procedure in which 28 items were presented to 70 students enrolled in a

Philosophy of Education course. Subjects read one of the three versions of each item and indicated whether the target should be interpreted idiomatically or literally, or whether the target was uninterpretable given the context preceding it. In addition, they indicated the degree to which a target followed from a particular context on a 7-point scale. Based on these ratings, 18 items were selected that were most consistently categorized and that were rated as being most closely related to their contexts. Eighteen additional items were constructed, and together with the nine rejected items in the norming study they appeared as practice items before the experimental ones.

Contexts ranged from 19 to 30 words in length with a mean length of 24 words (approximately half of the length of the long contexts in Experiment 1). The mean number of words was 24.3 for idiom and control contexts and 24.6 for literal contexts. The 18 idiomatic expressions ranged from 3 to 7 words with a mean of 4 words while the literal translations of these idioms, the control targets, ranged from 2 to 4 words with a mean of 3 words. For any particular item, the control target was always either the same length as or shorter than the idiomatic expression.

Three lists of experimental items were constructed. An equal number of idiomatic, literal, and control versions of items appeared in each list, with no repetition of items. For example, an item appearing in its idiomatic version in List A appeared as a control item in List B and as a literal item in List C. Two random orders of each list were constructed.

Procedure. Subjects were instructed exactly as in Experiment 1. They followed the same procedure with the following two exceptions. Half of the subjects completed the main phase of Experiment 1, received their payment, and then completed the main phase of this experiment. Also, all subjects in Experiment 2 received two types of questions on the posttest. The first type was essentially the same as in Experiment 1: the subjects were asked to write down as best they could remember how

they had interpreted the target phrases during the reaction time phase of the experiment. After this they again saw all of the idiomatic phrases and were asked to rate on a scale from 1 to 4 how well they understood the idiomatic meaning of these and how often they personally came across or used these expressions.

Results and Discussion

A first analysis involving a comparison of times spent reading the context segments of the items revealed no differences among groups, $p > .20$ in all cases. Before analyzing target times, the measure of prime interest, the data were examined and some data points were removed. These included reaction times that fell three standard deviations above a subject's own mean performance, and those which were below 500 milliseconds, amounting to a total of 30 cases, or 3.5% of the data. No more than one data point per subject was removed for these reasons. On one occasion, a subject did not read the context segment and her response on that target was discarded. Four-way analyses of variance were performed with list, order, and sequence as between subject variables and with context type as a within-subject variable. There was some evidence that the context type effect was significant, $F_1(2, 76) = 5.52, p < .01$, and $F_2(2, 34) = 3.01, p < .10$. Means indicated that idiomatic uses of idioms and the control targets (1472 and 1487 msec, respectively) were understood more quickly than literal uses of idioms (1682 msec).

Subjects' responses to questions on the posttest concerning their knowledge of the idioms were analyzed. The corresponding response times for those idioms which subjects indicated that they only vaguely understood or that they did not know at all were removed. These removals amounted to 5.5% of the data. When these response times were removed, as well as the very slow and very fast times removed in the first set of analyses (total of 9% of the raw data), results of the four-way

analysis of variance indicated a significant context type effect, $F_1(2, 74) = 9.02, p < .001$, $F_2(2, 34) = 3.66, p < .05$, $\min F'(2, 63) = 2.60, p < .10$, and $\max F'(2, 63) = 3.51, p < .05$. Newman-Keuls analyses revealed that idiomatic targets (1383 msec) and control targets (1486 msec) were understood significantly more quickly than literal targets (1677 msec), $p < .01$ and $p < .05$, respectively. Control and idiomatic targets did not differ from each other significantly.

Subjects' responses to the second posttest, the test of memory for target interpretations, indicated that in 78% of the cases, the subjects had interpreted the target as intended. "Forgot" responses occurred for 14% of the targets while 1% of cases could not be categorized as correct or incorrect. The 7% remaining represent the 37 instances in which targets meant to be interpreted literally were, correctly or incorrectly, reported by subjects to have been interpreted idiomatically. There were also three instances in which subjects reported a literal interpretation of idiomatic items. Analyses of variance, in which the response times corresponding to misinterpreted targets were deleted, failed to reveal conventionally significant results for the context type factor. However, mean response times for idiomatic, literal, and control items were ordered as in the analyses reported above, and a strong trend was indicated, $p < .15$. Analyses in which both posttests, the knowledge of idiom question and memory for target interpretation question, were used as a basis for removing data points once again indicated a significant context type effect, $F_1(2, 74) = 3.43, p < .05$.

In conclusion, the confirmation of the hypothesis that motivated Experiment 2 must be regarded as somewhat tentative. The data certainly demonstrate that idioms, a type of figurative language, do not take longer to comprehend than literal uses of those same expressions, and there are indications that they seem to be processed more quickly than literal language.

GENERAL DISCUSSION

For the most part, the results confirmed the hypotheses we set out to test. In Experiment 1 the targets requiring a metaphorical interpretation under conditions of minimal contextual support took longer to be understood than those requiring literal interpretations. This difference disappeared when the context length was increased. Experiment 2 revealed that familiar idioms are processed as quickly as, if not faster than, syntactically and semantically comparable literal language.

The most important finding from Experiment 1 is the interaction between context type and context length. The chief determinant of processing time was, as predicted, the degree of contextual support, and associated with that, the transparency of the relationship between a target and its preceding context. Our explanation of these results is that where there is little context the expectations that arise from it are insufficiently specific for the hypothesis/test process to be effective, and metaphors suffer significantly more than literals. Where there is an abundance of preceding context, the process is hardly less effective for metaphors than for literals. The interaction is also interesting when considered in the light of Clark and Lucy's (1975) findings. In the short context condition the present findings can be viewed as a replication of their results (applied to metaphors), but in the long context condition they cannot be so viewed. Our expectation, therefore, is that the Clark and Lucy findings would not replicate if the targets were presented under conditions wherein an antecedent context enabled subjects to generate correct expectations about the specific conveyed meaning of the indirect requests. Indeed, Rumelhart (in press) reports exactly such a study with the outcome being just as we would predict.

The account of the comprehension process that has been offered so far has relied heavily on an unexplained notion of "contextually generated expectations." Although somewhat

speculative in nature, a more precise account in terms of schema theory (see, e.g., Rumelhart and Ortony, 1977) will now be proposed. Consider first the long context condition for metaphors, taking the following context as an example:

Severe criticisms of Europe's oldest dictator came from within Spain and without. For 35 years Franco was barraged with these constant criticisms. To the end, the struggle continued between the ruthless dictator and his critics.

As a subject reads this passage, he invokes several schemata in order to comprehend it. They include those for *Franco, dictator, Spain, criticism, resistance, persistence*, and so on. Now the subject reads the target. *The waves beat relentlessly against the rugged coastline*, with these and related schemata already primed. The context has generated expectations that the input that follows it can be largely accounted for in terms of these schemata. And, indeed, it can, for the target can be accounted for by such schemata as those of *resistance* and *persistence* sufficiently well to permit those aspects of it that do not fit to be ignored. (Presumably readers set some criterion of what counts as a sufficient account of an input.) Thus, the notions of contextually generated expectations and "accounting for an input" are complementary.

The process underlying the comprehension of literals in the long context condition is, of course, very similar. Using the same item for an example, the subject sees:

Iceland's coastal region was windswept. The angry arctic winds often buffeted the fishing villages located on the numerous bays and inlets. The darkening skies and the rising winds announced the onset of another storm.

The schemata employed in the comprehension of this context presumably include those for *Iceland, coast, villages, bays, storm, wind, sea, waves*, and associated with these, schemata for *resistance* and *persistence* again. Thus, it would appear that most of the schemata required to account for the target would be

available, just as they were for the metaphor, and in this case at least, perhaps one or two more than for the metaphor. So, in both long context conditions it seems that processing the context activates sufficient appropriate schemata to enable an account of the target to be arrived at quite readily, even though there may be slightly more of them available for literals than for metaphors.

In the short context conditions, things are rather different. Using the same example, for the literals the context is comprised of only the first sentence of the long context, *Iceland's coastal region was windswept*. Certainly fewer appropriate schemata can be activated in the comprehension of this context, and consequently the expectations that can be generated from it are very vague compared to those in the long context condition. Nevertheless, many of the schemata that are available will help. In the short context condition for metaphors this is not true, for here all the subject sees is, *Severe criticisms of Europe's oldest dictator*. There can be almost no schemata resulting from the comprehension of this statement that would immediately permit a satisfactory account of the target to be given. The very slow times found for the short context conditions might be attributed to a second, related factor. One could regard the subjects as having to engage in a sort of problem solving. They are given two pieces of information (context and target) and before they respond they have to determine some reasonable relationship between them. It would be neither surprising nor unreasonable to suppose that under these conditions a subject might delay generating a more elaborated representation of the context until *after* seeing the target. Such a strategy—the use of a sentence to clarify and to elaborate the interpretation of a preceding one—is surely quite common in normal reading and is very likely to be reflected in an increase in processing time.

We have argued that when the context is read a number of schemata are activated, at least some of which can be used to account for

the target. Our speculations about the relative numbers of appropriate schemata activated by the contexts in the different conditions seem to accord rather well with the ordering of reaction times in those conditions. In the long context condition most of the schemata employed in processing the context can be used immediately to account almost completely for the literal target. In the long context condition, again, many of them, but perhaps not quite so many, can be employed to give a satisfactory account of the metaphorical target. In the short context condition, not many schemata are available. Those that are, turn out to be very helpful for literal targets, but almost none of them are immediately helpful in accounting for metaphorical targets. If an account of this nature is to be accepted, it is, of course, necessary to assume that there is a close connection between the availability of appropriate schemata in terms of which a target can be interpreted and the time taken to make the interpretation. This assumption, however, is a perfectly reasonable one. When insufficient schemata are available it is to be expected that a subject will have to introduce schemata not directly activated by the context. In some cases, as in the short context, literal condition, this may be very easy if it happens at all. In other cases, as in the short context, metaphorical condition, the schemata needed may be semantically relatively remote. In both cases, the scarcity of appropriate schemata, coupled with an abundance of irrelevant ones, is going to increase the difficulty of accounting for the input and consequently the time required to do so.

The position that we are advocating suggests that, in general, figurative language is processed in much the same way as is literal language. What determines the difficulty of processing is not nonliteralness but relatedness to context. We have argued that relatedness to context can be high or low for both literal and nonliteral uses of language. Evidence for this was found in Experiments 1 and 2. Although one cannot infer from Experiment 2 that

idioms are processed faster than all comparable literal expressions, it seems to be the case that idiomatic uses of idiomatic expressions are processed faster than literal uses of them, which are comparable. This difference is probably due to one of two reasons, or to some combination of both. First, if one assumes that the relatedness of target to context was comparable in all conditions, then the familiarity of the idioms may have led subjects to try to interpret them idiomatically before trying to interpret them literally. Evidence that subjects will take longer to respond to the nonpreferred interpretation of an ambiguous expression has been found (e.g., Foss, Bever, & Silver, 1968). The second reason relates to the possibility that the meaning of an idiom may be stored in much the same way as the meaning of a single lexical item. This would result in much lower processing demands at the syntactic level than a non-idiomatic expression of comparable syntactic structure. In this connection it is interesting to note that the idioms were greater in length than the control expressions (e.g., *let the cat out of the bag* is longer than *reveal the secret*) yet the mean times for them were, if anything, shorter than those for the control expressions. Accordingly, it is tempting to conclude that idioms are actually processed significantly faster than unambiguous expressions of similar length. If this is indeed the case, it is probably due to both of the two factors just described. Conventional uses of idioms would have a parsing advantage over comparable literal expressions, while the literal use of an idiomatic expression would be slowed by the tendency to automatically go for the idiomatic meaning. However, the present experiment, while suggestive, is not conclusive on this issue.

It might be argued that Experiment 2 should have employed idioms whose interpretations were as likely to be the literal ones as the idiomatic ones. In this way, the disadvantage suffered by the less preferred meaning would have been eliminated. However, this would

have necessitated using much less familiar idioms and subjects who were not familiar with them would, for the most part, have contributed nothing to the data. This is because most idioms are not "frozen" metaphors. If the idiomatic meaning is not known, it cannot, as a rule, be figured out, as can the meaning of novel metaphors. Idioms tend to be instances of frozen metonymy and are based on often highly specialized local customs or habits. For example, the idiom "kick the bucket" derives its origins from an old, and now obsolete, practice in the South of England. When pigs were taken to be slaughtered, they were traditionally tied up in such a way that their back legs would be constantly kicking a beam called a "bucket." Thus, the relationship with dying is a metonymical one, and one that could not be determined by engaging in those processes that are usually employed in the comprehension of novel metaphors, unless one were endowed with this particular piece of esoteric, and, for the most part, useless knowledge! Thus, in general, the comprehension of unfamiliar idioms cannot be achieved by using a semantic analysis of the expression as can novel metaphors, so the stage model could not possibly apply. People usually have to learn what an idiom means, not figure it out. If they have learned it and the idiom is a familiar one, then its conventional meaning is directly associated with it. If they have not learned it and it is an unfamiliar one, then very often its conventional meaning cannot be determined from its literal meaning, although, of course, it might be determinable from the context alone.

The results of these two experiments lead us to conclude that the account we have offered is reasonable. In the normal course of events nonliteral uses of language, be they metaphors, idioms, or indirect speech acts, are comprehended without any special processing. The predominantly top-down strategy that is employed in language comprehension enables them to be comprehended in terms of the preceding context quite naturally. On the other

hand, there certainly are cases where an utterance is insufficiently related to the context for it to be understood. These cases include literal as well as nonliteral uses of language. In such cases, it is necessary to engage in additional inferential procedures, and these seem to be well captured by the stage model. However, the stage model says nothing about the nature of the reinterpretation stage. We suggested that the determination of the ground of the metaphor might represent one such strategy. This should not be taken to imply that the ground of a metaphor is not determined when there is adequate contextual support. In unproblematic cases the determination of the ground is no more and no less than the utilization of those activated schemata required to account for the metaphor. That is, in cases where comprehension proceeds smoothly, the ground does not have to be "computed." However, in cases where the stage model is the appropriate description, it does. A reader or listener has to make inferences based on the available schemata, and on expectations and knowledge about the speaker or writer, and the speaker's or writer's intentions. This is done quite consciously when one tries to make sense of obscure poems and other works of literature. It is perhaps done less consciously in more mundane cases, but it is done nevertheless.

REFERENCES

- BRANSFORD, J. D., & JOHNSON, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 717-726.
- CLARK, H. H., & LUCY, P. Understanding what is said from what is meant: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, 1975, 14, 56-72.
- FOSS, D. J., BEVER, T. G., & SILVER, M. The comprehension and verification of ambiguous sentences. *Perception and Psychophysics*, 1968, 4, 304-306.
- GRICE, H. P. Logic and conversation. In P. Cole & J. L. Morgan (eds.), *Syntax and semantics: Speech acts*. New York: Academic Press, 1975. Vol. 3.

- HARRIS, R. J. Comprehension of metaphor: A test of a two stage processing model. *Bulletin of the Psychonomic Society*, 1976, **8**, 321-324.
- HAVILAND, S. E., & CLARK, H. H. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 512-521.
- ORTONY, A., REYNOLDS, R. E., & ARTER, J. Metaphor: Theoretical and empirical research. *Psychological Bulletin*, 1978, **18**, 919-943.
- RUMELHART, D. F. Some problems with the notion of literal meanings. To appear in A. Ortony (ed.), *Metaphor and thought*. London/New York: England: Cambridge Univ. Press, in press.
- RUMELHART, D. E., & ORTONY, A. The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N. J.: Erlbaum, 1977. Pp. 99-135.
- SCHALLERT, D. L. Improving memory for prose: The relationship between depth of processing and context. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 621-632.
- SEARLE, J. R. *Speech acts*. London/New York: England: Cambridge Univ. Press, 1969.
- SEARLE, J. R. Metaphor. To appear in A. Ortony (ed.), *Metaphor and thought*. London/New York: England: Cambridge Univ. Press, in press.
- STERNBERG, R. J. Component processes in analogical reasoning. *Psychological Review*, 1977, **84**, 353-378.
- TVERSKY, A. Features of similarity. *Psychological Review*, 1977, **84**, 327-352.
- VERBRUGGE, R. R. Resemblances in language and perception. In R. Shaw and J. Bransford (eds.), *Perceiving, acting and knowing*. Hillsdale, N. J.: Erlbaum, 1977. Pp. 365-392.
- VERBRUGGE, R. R., & MCCARRELL, N. S. Metaphoric comprehension: Studies in reminding and resembling. *Cognitive Psychology*, 1977, **9**, 494-533.

(Received October 31, 1977)