

The effectiveness of context before, after, and around a missing word

DAVID C. RUBIN

Lawrence University, Appleton, Wisconsin 54911

The probability of correctly guessing a missing word was measured using four different kinds of context: all words before the missing word (forward context), all words after the missing word (backward context), all words before and the one word after the missing word (surround context), and just the one word after the missing word. The probability of correctly guessing a missing word was greater with the forward than with the backward context. The probability of guessing correctly with the surround context was much greater than would be predicted from the independent combination of its forward and one word after components. The results provide evidence that expectations are formed continuously during comprehension, but not in a strict word-by-word order. Implications for information theory are noted.

Expectations are known to be important in the normal comprehension of language (e.g., Neisser, 1967). This paper investigates expectations arising from context occurring before, after, and around a missing word. By completely removing the stimulus word from its context, expectations can be studied without attempting to study the interaction of those expectations with the perception of the physical stimulus itself.

Language is normally presented in a fixed temporal order, that is, from left to right. It is therefore reasonable to assume that through practice, or the evolved structure of language, context to the left of a word would come to produce more accurate expectations than context to the right of that word. While this assumption has considerable support (Bever, 1970; Forster, 1974; Garrett, 1974; Howes, 1974), it has been neither directly tested nor universally accepted. For instance, transformational grammars, as well as some nontransformational grammars (Kay, 1973) do not begin processing a sentence until after it is complete. Such grammars should not favor context to the left over context to the right, as all context is available and necessary at the time the sentence is parsed. Similarly, information theory predicts equivalent redundancy given context to the left and right of a missing word.

Even if language is processed in a general left-to-right order, transient memories could allow judgments to be postponed for a period of time longer than the duration of individual words. That is, expectations need not be formed strictly on a word-by-word basis. Rather, a small amount of context to the right of a word could be used to limit the possible alternatives. If efficient use is normally made of such context in comprehension, then adding

a small amount of context to the right of a missing word should greatly increase its predictability.

The distinction between a word-by-word and a wait-and-see strategy has implications for current theories of comprehension (Garrett, 1974). For instance, Kaplan and Wanner have proposed a formal computer model of comprehension which accounts for many, otherwise unsystematic, psychological findings (Kaplan, 1975; Wanner & Maratsos, Note 1). Using a strict word-by-word, left-to-right order, the model continuously guesses at the syntactic structure of a sentence as each word in that sentence is presented. A large effect of postcontext might suggest a modification to delay judgment until more context is available.

The following experiments are an attempt to resolve some of these issues empirically. While expectations have been studied extensively (Garner, 1962; Rankin, 1965), only two kinds of context have generally been used. Either the context has consisted of words prior to the missing word (forward context) or else the context has been a passage from which several words have been deleted (e.g., Miller & Coleman, 1967). In Experiment I, three types of context were studied: (1) the standard forward context, (2) backward context which consists of the words after the missing word, and (3) surround context which consists of forward context plus the one word which follows the missing word. A between-subjects design was used in which each subject guessed at words missing from a passage given only one of these three types of context.

EXPERIMENT I

Method

Subjects. Thirty-six undergraduates volunteered for the experiment.

Materials. Three 150-word passages were used. The passages were calibrated by Miller and Coleman (1967) as having nearly equivalent forward context predictability (34.3%, 34.4%, and

I am indebted to Francis Campos, Richard Long, Donald Olivier, and Eric Wanner for their assistance.

34.4% correct). Fourteen words from each passage (every 10th word except the last word) were chosen as missing words.

Booklets were used to present the stimulus material. The first page of the forward context booklet consisted of the first 9 words of a story and a blank space. The second page consisted of the first 19 words of that story and a blank space. The third page consisted of the first 29 words of the story, and so forth until the 15th page, which contained the complete story with no blanks.

The format for the surround context consisted of the forward contexts plus the one word which followed the missing word.

The first page of the backward context consisted of a blank space followed by the last 10 words of a passage. The second page consisted of a blank followed by the last 20 words of that story, and so forth until the story was completed.

The combination of the three contexts with the three stories resulted in nine unique context-story pairs. Four booklets were prepared for each of these nine pairs.

Procedure. Each of the 36 subjects was given a booklet and instructed to fill in each blank with the word they thought was most likely to have been in the original text. Examples of the three types of context were given in the instructions. The task was self-paced. Following the traditional procedure (Miller & Coleman, 1967), a response was scored correct only if it was identical to the missing word. Misspellings were allowed, but only if they did not result in a change of tense or number.

Results

The effect of the context was significant [$F(2,27) = 18.69, p < .0001$]. The effects of the story [$F(2,27) = 3.06, p = .06$] and the Context by Story interaction [$F(4,27) = 0.20, p = .94$] were not. The means for the context main effect were 36.3% ($SD = 11.6\%$), 22.6% ($SD = 11.3\%$), and 50.6% ($SD = 11.2\%$) for the forward, backward, and surround contexts, respectively. The differences between the means were significant at the .01 level (Newman-Keuls test).

The marginally significant main effect of stories, which indicates only that the stories did not have equally easy missing words, is of little importance here, as the Story by Context interaction is nonsignificant. In particular, the forward, backward, and surround context means for the three stories considered individually are: 41.1%, 21.4%, 50.0% [$F(2,9) = 8.04, p = .01$]; 30.4%, 25.0%, 51.8% [$F(2,9) = 6.92, p = .02$]; 37.5%, 21.4%, 50.0% [$F(2,9) = 3.74, p = .06$]. Thus, the effect of context is the same for the three stories sampled. A more severe test of the generality of the results is to consider both missing words and subjects as random variables. Using Clark's (1973) formulas 15 and 16 to calculate the minimum quasi F ratio, the effect of context is again significant [$\min F'(2,103) = 6.05, p < .005$] while the effect of story [$\min F'(2,55) = .56$] and the Story by Context interaction [$\min F'(4,103) = .06$] are not.

It should be noted that the passages and missing words were sampled without regard to semantic or syntactic structure. It is quite possible that sampling in a semantically or syntactically determined manner could lead to different results. For instance, choosing passages with plots that only become clear in the last sentence or choosing the missing words to always be

the first words of a sentence might reduce the superiority of the forward over the backward context. Likewise, choosing missing words to always be the last word of a clause or sentence might reduce the superiority of the surround over the forward context.

Subjects using the surround context consistently had one more word of context than subjects using the forward context: that is, they had an average of 75 as opposed to 74 words of context. The difference in the number of words of context provided, however, cannot account for the superiority of the surround context. First, there was no statistically significant increase in the probability of a correct guess as a function of position in the passage. Second, the forward context could have been given an average of 9 more words of context than the surround context by eliminating from the scoring the first missing word of the forward context and the last missing word of the surround context. This deletion of blanks would have left an average of 79 words of context in the forward direction as opposed to 70 in the surround. The difference between the surround and forward context would still, however, have remained significant [35.1% vs. 48.8%, $t(22) = 2.83, p < .01$].

EXPERIMENT II

The surround context consists of the forward context plus the one word after the missing word. In order to examine the relation of the forward and surround contexts in more detail, the probability of guessing correctly given only the one word after the missing word was measured.

Method

Subjects. Twelve undergraduates, who did not take part in Experiment I, volunteered.

Materials. For each of the 42 missing words from Experiment I, a 4 x 6 in. index card was prepared. On it was typed a code number and a blank space followed by the one word after the missing word.

Procedure. Subjects were run individually. Each subject was asked to fill in every blank in the 42 shuffled cards. The subjects were informed that these contexts came from several different passages.

Results

The average probability of guessing a missing word correctly given only the one word after was 0.048 ($SD = 0.020$).

Given this probability, the relation of forward and surround context can be examined in more detail. Let the probability of guessing correctly with the forward context be P_f and the probability of guessing correctly with the one-word-after context be P_a . Assume these two sources of information are independent, that is, if P_f is unaffected by the presence of the one-word-after context and P_a is unaffected by the forward context. Then, with both sources of information, the probability of guessing correctly, $P_{\text{independent}}$,

would be $P_f + P_a - P_f \cdot P_a$ (Pollack, 1964; Tulving, Mandier, & Bauml, 1964). In other words, if subjects used the information from the forward context and information from the one-word-after context independently, then the probability of guessing, given both contexts (i.e., given the surround context), would be $P_f + P_a - P_f \cdot P_a$ or 0.394. This probability is well below that actually observed for the surround context. [Assuming the most conservative case of $SD^2_{independent} = SD^2_f + SD^2_a$, $t(22) = 2.29$, $p < .05$].

Thus, the two sources of information are not used independently, but are used in conjunction to limit the possible guesses.

DISCUSSION

The probability of guessing a missing word correctly is greater given the n words before it than given the n words after it. This overall left-to-right advantage, however, does not hold on a word-to-word level. Over this short range of context, providing one additional word after a missing word provides much greater constraint than providing 10 additional words before it. These results can be accounted for in terms of a user-language system which has developed to operate on temporally ordered stimuli. Over the short range, however, the stimuli need not be processed in their original temporal order, as sensory information stores and short-term memory can be used to hold segments of the stimuli longer than one word.

The study also has implications for information theory. According to information theory, the forward and backward contexts should have yielded the same redundancy (see Garner, 1962, p. 225, or Shannon, 1951, p. 58, for formal proofs). In fact, Shannon (1951) has demonstrated this equality for missing letters and Garner (1962) has gone as far as to argue that the 50% forward redundancy of letters in English exists just because 50% redundancy in the forward direction implies 50% redundancy in the backward direction, yielding a desirable 100% bilateral redundancy. While this paper remains mute on letter redundancy, it does demonstrate that for words the probability of guessing correctly is greater with forward than with backward context. The actual redundancy measure considered by information theory was not used here because of the extremely large number of subjects that would have been needed for reliable calculations. It is therefore necessary to assume that, over the sample of contexts presented, the greater predictability of the forward context implies a greater redundancy. As an examination of the data indicates that subjects did not make more of the same incorrect guesses in the backward than in the forward context (which would have been necessary

to increase redundancy without increasing predictability), this assumption seems reasonable. The results therefore indicate that the common use of human subjects to obtain estimates of redundancy must be reexamined. At least at the word level, the systematic differences between forward and backward context indicate that, while estimates of information based on human judgment may provide valid measures of information in the general sense of the word, they cannot provide a valid measure of information in the statistical sense.

REFERENCE NOTE

1 Wanner, E., & Maratsos, M. *An augmented transition network model of relative clause comprehension*. Unpublished manuscript, Harvard University, 1975.

REFERENCES

- BEVER, T. G. The cognitive basis for linguistic structures. In J. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley, 1970.
- CLARK, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 335-359.
- FORSTER, K. I. Linguistic structure and sentence production. In C. Cherry (Ed.), *Pragmatic aspects of human communication*. Dordrecht, Holland: Reidel, 1974.
- GARNER, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.
- GARRETT, M. Experimental issues in sentence comprehension: Complexity and segmentation. In C. Cherry (Ed.), *Pragmatic aspects of human communication*. Dordrecht, Holland: Reidel, 1974.
- HOWES, D. The link between speech production and speech perception. In H. R. Moskowitz & B. Scharf (Eds.), *Sensation and measurement Papers in honor of S. S. Stevens*. Dordrecht, Holland: Reidel, 1974.
- KAPLAN, R. On process models for sentence analysis. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1975.
- KAY, M. The MIND system. In R. Rustin (Ed.), *Natural language processing*. New York: Algorithmics Press, 1973.
- MILLER, G. R., & COLEMAN, E. B. A set of thirty-six prose passages calibrated for complexity. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6, 851-854.
- NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
- POLLACK, I. Interaction between auditory and visual information sources in word identification. *Language and Speech*, 1964, 7, 76-83.
- RANKIN, E. A. The cloze procedure—a survey of research. In E. L. Thorstone & L. E. Hafner (Eds.), *The philosophical and sociological bases of reading, Fourteenth yearbook of the National Reading Conference*. Milwaukee, Wis: National Reading Conference, 1965.
- SHANNON, C. E. Prediction and entropy of printed English. *The Bell System Technical Journal*, 1951, 30, 50-64.
- TULVING, E., MANDLER, G., & BAUMAL, R. Interaction of two sources of information in tachistoscopic word recognition. *Canadian Journal of Psychology*, 1964, 18, 62-71.

(Received for publication April 9, 1975;
revision accepted December 2, 1975.)