COGNITIVE PSYCHOLOGY 5, 257–274 (1973)

Reading Rate and Retention as a Function of the Number of Propositions in the Base Structure of Sentences[1]

WALTER KINTSCH AND JANICE KEENAN
University of Colorado

English texts were constructed from propositional bases. One set of 16-word sentences was obtained from semantic bases containing from 4 to 9 propositions. For another set of sentences and paragraphs, number of words and number of propositions covaried. Subjects read the texts at their own rate and recalled them immediately. For the 16-word sentences, subjects needed 1.5 sec additional reading time to process each proposition. For longer texts, this value increased. In another experimental condition reading time was controlled by the experimenter. The analysis of both the text and the recall protocols in terms of number of propositions lent support to the notion that propositions are a basic unit of memory for text. However, evidence was also obtained that while the total number of propositions upon which a text was based proved to be an effective psychological variable, all propositions were not equally difficult to remember: superordinate propositions were recalled better than propositions which were stucturally subordinate.

Considerable experimental evidence exists today indicating that sentences, and prose material in general, are not stored in memory verbatim, but are coded as to their content (Sachs, 1967; Bransford & Franks, 1971; Kintsch & Monk, 1972). While such findings agree very well with our intuitions, they do not help us determine precisely in what way content is coded in memory. Since the general problem is extremely complex, no general theory of text memory has yet been formulated. Such a theory would be required to consider numerous properties of a text. For example, a sentence presents some new information in a given context, situational or verbal; a part of a sentence or paragraph must be identified as the theme; there are grammatical relations among the elements of a sentence; and finally, each text has a logical structure. It is only this latter, ideational function of language that the present paper shall be

[1] This research was supported by a Grant from the National Institute of Mental Health, MH-15872. The research findings reported in this paper were presented at the Meeting of the Psychonomic Society, St. Louis, 1972. Reprint requests should be sent to: Dr. Walter Kintsch, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

concerned with. Textual and communicative aspects will be neglected, even though they are undoubtedly of great psycholinguistic significance.

A number of theories have recently been developed concerning the logical structure of text and, closely related, the nature of semantic memory. Several linguists have explored the idea that sentences have a base structure consisting of propositions which represent their semantic content. Kintsch (1972) has shown how one could conceive of semantic memory as being constructed from such propositional expressions. In the present paper an attempt is made to show that the propositional base structure of a sentence, as postulated by Kintsch (1972), has predictable empirical effects upon behavior. Such a demonstration is of crucial importance if the speculations that have been offered about semantic memory are to be taken seriously as a psychological model.

This initial experiment was designed to determine whether the number of propositions in the base structure of a sentence is an important psychological variable. In particular, sentences which differed in the number of underlying propositions, but which were equal in terms of total number of words, were given to subjects to read and to recall immediately after reading. A second set of sentences and paragraphs, in which number of words and number of propositions covaried, was also used in order to study number of propositions over a wider range. With the number of words per sentence controlled, propositions varied from 4 to 9; with the second set of sentences propositions varied from 2 to 23. If number of propositions is an effective psychological variable, reading time should increase as more propositions are processed; likewise, the number of propositions subjects are able to recall should be related to processing time in a lawful way. Of course, by averaging over all kinds of different propositions and looking only at the effect of the total number of propositions, many potentially significant sources of variance are being neglected. In addition to the mere number of propositions, the nature and number of their arguments, their complexity, structural relations among them, lexical factors, and many other properties might be important determinants of reading speed and comprehensibility. In the studies reported here, all of these factors vary unsystematically. Thus, the difficulty that each proposition presents is certainly not a constant, but a random variable with an appreciable variance. Nevertheless, as long as one deals with averages, the number of propositions upon which a text is based should have a strong and significant effect, if the proposition is indeed a proper unit of analysis for studies concerned with the processing and storage of linguistic information.

Some subjects were allowed to study each sentence or paragraph as long as they wanted; their reading time was one of the major dependent variables of interest for the present study. Other subjects, however, were given only a limited reading time. Their recall was compared with that of unrestricted subjects, and a processing model that captures some of the most salient features of both sets of data was developed.

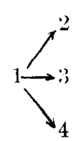
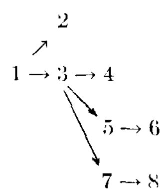
METHOD

Subjects. In the free reading time condition, 29 undergraduates from the University of Colorado served as subjects. They were fulfilling part of a course requirement. For the restricted reading time condition, 44 students were each paid $2.00 for their participation.

Material. Two sets of materials were constructed for this experiment. Set A consisted of 10 sentences which were 16–17 words long counting punctuation, 14–16 otherwise. The sentences were not related to each other, most of them dealing with topics from classical history. This choice of topics was made in an attempt to hold the relative familarity of the text at a minimum, while avoiding problems of vocabulary concomitant with equally unfamiliar but more technical material.

Although word length in these sentences was fairly strictly controlled, the number of propositions upon which each sentence was based varied between four and nine. Two sample sentences, together with the propositions from which they were constructed and the hierarchical relationships which exist among these propositions, are shown in Table 1. The analysis into propositions was made according to the theory described in Kintsch

TABLE 1
Two Sample Sentences from Set A (Sentences I and VIII), Together with Their Propositional Analyses

I.	Romulus, the legendary founder of Rome, took the women of the Sabine by force.	
1	(TOOK, ROMULUS, WOMEN, BY FORCE)	
2	(FOUND, ROMULUS, ROME)	
3	(LEGENDARY, ROMULUS)	
4	(SABINE, WOMEN)	
VIII.	Cleopatra's downfall lay in her foolish trust in the fickle political figures of the Roman world.	
1	(BECAUSE, α, β)	
2	(FELL DOWN, CLEOPATRA) = α	
3	(TRUST, CLEOPATRA, FIGURES) = β	
4	(FOOLISH, TRUST)	
5	(FICKLE, FIGURES)	
6	(POLITICAL, FIGURES)	
7	(PART OF, FIGURES, WORLD)	
8	(ROMAN, WORLD)	

(1972). Briefly, each proposition consists of a relational term (written first) and one or more arguments. Propositions may themselves appear as arguments of other propositions, in which case Greek letters were used in order to avoid overly complex expressions. The analysis used here differs from that proposed earlier (Kintsch, 1972) in only one minor respect: the restriction that verbs may not be used as arguments in propositions has been dropped; this led to some considerable simplifications in the obtained expressions. Propositions have been numbered solely for the purpose of identification in the graphs showing the hierarchical relations among them. As described in Kintsch (1972), these hierarchical relations are implicit in the order in which propositions are written: a proposition β is subordinated to another proposition α if α precedes β in the list of propositions and if α and β have at least one term (relation or argument) in common.

One of the major problems in work of this type is that no algorithmic procedure exists to analyze a given sentence (or paragraph) into its propositional base structure. However, for present purposes the problem can be circumvented in the following way: one can start with the propositional expressions themselves and translate these into English text. The rules for doing so are simpler and are somewhat better understood, although they do not really exist in explicit form either. At least very substantial agreement could be obtained that the sentences used are indeed one way to express the meaning of the coordinated propositions.

Set B consisted of 20 sentences or paragraphs in which the number of underlying propositions and the number of words were confounded. The shortest sentence contained seven words in two propositions; the longest paragraphs were based upon 22 propositions, requiring 58 words, and 23 propositions, requiring 45 words. All sentences in Set B were definitions and descriptions of psychological terms, modified from the glossaries of several psychological texts, and thus, dealt with relatively unfamiliar topics.

Procedure—free reading. Slides were made for all sentences and paragraphs. The subject was seated about 3 m in front of a screen on which the slides were shown by means of a Kodak Carousel projector. The experimenter, who was seated behind the subject, exposed a slide and simultaneously started a Hunter Timer. In front of the subject there was a box with two labeled response buttons: an Advance slide button and a Finish button. The subject was instructed to read the sentence carefully and to press the Advance button in front of him when he was finished. This response stopped the timer, started a second timer, and removed the sentence from the screen. The subject then recalled the sentence he had just seen in writing. Each sentence was written on a separate page of a booklet. When the subject finished writing, he pressed the Finish button which stopped the second timer and turned on an indicator light at the experimenter's console. After recording the results, the experimenter made sure that the subject had turned to a new page and had put his finger on the Advance button; the experimenter then gave the subject a "Ready" signal and started the next trial.

One warm-up problem preceded the 30 experimental trials. New random orders of the experimental slides were constructed for every three subjects. The instructions emphasized that exact wording was not as important as the meaning of the sentences, and the subjects were asked to work as fast as possible, keeping in mind that the important thing was how much they remembered, not speed.

Restricted reading time. The procedure was identical to that used in the free-reading time condition, except that the sentence slide was exposed only for a certain predetermined amount of time, and the subjects disregarded the Advance button. The exposure time for each sentence was made proportional to its length in terms of the number of words. In the free reading condition, the overall average reading time per word was .97 sec. A third of that, .33 sec, was taken as the exposure time per word for the restricted reading time condition. The choice of one-third was made because it had been observed in earlier work (Kintsch & Monk, 1972) that restrictions of that magnitude still allowed the subjects to read through a text under normal circumstances, but were severe enough to produce noticeable behavioral effects. Thus, each sentence of Set A was exposed for 5 sec. For Set B, exposure times increased from 2 sec for the shortest sentences to 19 sec for the longest paragraph.

RESULTS

Sentence Set A. The subjects' recall protocols for the 10 sentences in Set A were scored for propositional recall. For each protocol it was determined which of the propositions were recalled. Paraphrases of the original wording were accepted as correct, as long as the propositional meaning was accurately expressed. If a subject made an error in a superordinate proposition which then reappeared in a subordinate proposition, which was otherwise correct, the subordinate proposition was accepted as correctly recalled, while the superordinate proposition itself was scored as incorrect. For example, suppose a subject recalled Sentence I as *Romulus took the Sabine cities by force*. Proposition 1 was scored as incorrect because of the substitution of *cities* for *women*, but Proposition 4 was scored as correct. Although the subject had made an error in the superordinate proposition, he had recalled correctly that, whatever the superordinate proposition was concerned with, it was *Sabine*. Scoring was done independently by the two authors who agreed in 95.2% of all recall protocols. All disagreements, most of which were simply errors by one or the other scorer, could be resolved in conference.

Mean reading times as a function of the number of propositions in the base structure of the sentences are shown in Fig. 1. A least square line,

$$t = 6.37 + .94P_{\text{pres}} \tag{1}$$

is also shown. For these sentences, which were all of the same length, the subjects took about one additional second reading time per proposition. The subjects did not, however, always recall a sentence perfectly. Since there probably was very little, if any, forgetting with the brief sentences in Set A, one can assume therefore that the subjects did not always process all the propositions in a sentence as they were presented. Thus, a better estimate of the reading time needed by subjects to process propositions can be obtained if reading time is computed as a function of the number of propositions that the subjects were actually able to recall on each trial, irrespective of how many propositions a sentence contained. These data are also shown in Fig. 1. It is clear that if propositions recalled rather than propositions presented are considered, the dependence of reading time upon number of propositions in a sentence

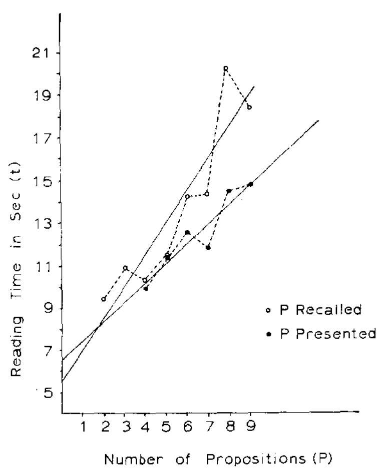


FIG. 1. Mean reading time for the sentences of Set A as a function of the number of propositions in the base structure of the sentences as presented, and as a function of the number of propositions recalled by the subjects, together with least square lines. The standard error for the points shown ranges from .75 to 2.10 sec, since points are based on differing numbers of observations, and averages 1.15 sec.

is even more pronounced. The least square line in this case is

$$t = 5.53 + 1.48P_{rec} \tag{2}$$

which means that every proposition processed required about 1.5 sec additional reading time. An analysis of variance confirmed this effect to be statistically reliable. The average reading times for each subject were computed, providing the subject recalled 2/3, 4, 5, 6, 7, or 8/9 propositions. (The pooling was done to avoid too many empty cells at the extremes.) The F value for Propositions was $F(5,112) = 10.9$, $p \leq .01$ (the degrees of freedom for the error term take into account some missing cell entries). A trend analysis reavealed most of this effect to be due to deviations from linear trend (87% of the variance associated with propositions, which was highly significant), while quadratic and higher order trends resulted in F values less than one. Estimating the proportion of the total within-subject variance of the reading times accounted for by linear regression upon the number of propositions recalled produced est. $\omega^2 = .21$.

Mean reading times were also computed separately for those instances when the subjects recalled a complete sentence correctly, and when they recalled only some of the propositions of a sentence. There was no systematic difference between these two sets of data, which justifies lumping them together as was done in Fig. 1.

Figure 2 allows a comparison between recall after free reading and

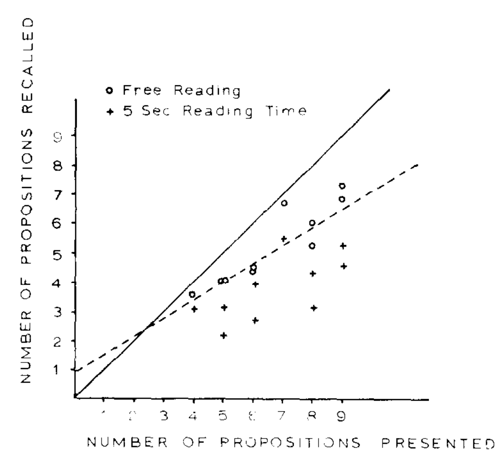


FIG. 2. Mean number of propositions presented and recalled for the sentences of Set A. The predictions for the free reading data are shown by the broken line (for explanation see text). The standard error for free reading is .22 propositions and .17 propositions for the 5 sec reading time.

limited reading time. The free reading data are partly implied by the results already discussed: Equations (1) and (2) can be combined to obtain a relationship between the number of propositions presented and recalled. This relationship is shown as the broken line in Fig. 2. Obviously, it describes the actual data quite well, although its slope (.64) somewhat underestimates the least squares value (.69). Number of propositions presented and mean number recalled correlate $r = .91$ in the data shown. This high value is not quite matched by the data from the restricted reading condition where an $r = .74$ was obtained. When reading time was limited recall was not as good as when reading time was free, and this difference was greatest for the most difficult sentences, that is, those based upon a large number of propositions. However, the same kind of relationship between propositions presented and propositions recalled appears to hold for both conditions.

When reading was self-paced, the subjects recalled 80% of the propositions correctly. Given the design of the present experiment, one cannot, of course, tell the cause of errors in this experiment: it may be a failure of processing, or forgetting, or a combination of both. Therefore, only correct responses will be discussed here. The percentage of propositions recalled of those actually presented was independent of either the total number of propositions in each sentence, the total number of terms appearing in the propositions for each sentence, or the reading time for that sentence. However, which propositions were recalled was by no means random. The hierarchical relationships among the propositions in each sentence were a powerful determinant of recall. There are two obvious ways to quantify these hierachies. One is to consider the rank of each proposition in a sentence, with the most superordinate proposition assigned rank 0, the immediately subordinate propositions rank 1, etc. Thus, in Sentence VIII of Table 1, Proposition 1 would have rank 0; Propositions 2 and 3 would have rank 1; Propositions 4, 5, and 7 would have rank 2; and Propositions 6 and 8 would be assigned rank 3. The likelihood of an error was computed for all propositions as a function of rank thus defined and is shown in Fig. 3. Of course, propositions of high rank must necessarily come from sentences with many propositions. In order to avoid possible selection effects, the propositions of each sentence were divided into two classes—low and high rank. When rank was Vincentized in this way, the decrease in the likelihood of a correct recall was still correlated with rank: recall was 86% for the superordinate propositions and 74% for the subordinate propositions.

A second way of quantifying propositional hierarchies is to count the number of descendants for each proposition. For example, in Sentence VIII, Proposition 1 has 7 descendants, Proposition 3 has 5 descendants,

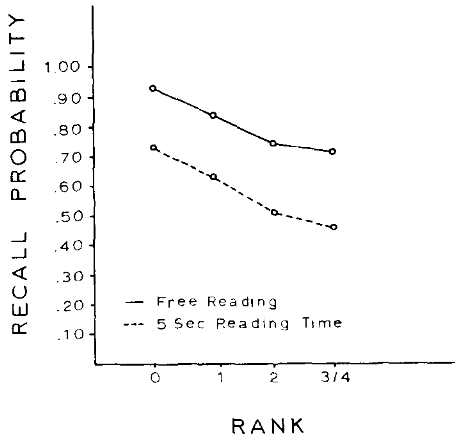


FIG. 3. Recall probabilities of propositions as a function of their rank in the sentence hierarchy, for both free and paced reading time. Ranks 3 and 4 are pooled because too few rank 4 observations were available. N values for the various points range from 290 to 1100.

Propositions 5 and 7 have 1 descendant each, while Propositions, 2, 4, 6, and 8 have no descendants. Calculating recall as a function of number of descendants also reveals a strong tendency for propositions with fewer descendants to be recalled more poorly. Propositions with 0–2 descendants were recalled 76% of the time, while propositions with 3 or more descendants were recalled 90% of the time. The two analyses in terms of rank and descendants were combined by computing separately recall of propositions as a function of their rank for propositions with and without descendants (about half of all propositions did not have descendants). While the probability of recall was strongly influenced by rank, the number of descendants had no effect upon recall if rank was controlled.

Several other factors were found to be important determinants of the recall likelihood of a proposition. For instance, many propositions appear in the surface structure as adjectival or adverbial modifiers. The error rate for those propositions is an astounding 30%. On the other hand, proper names were used as modifiers in the experimental sentences quite frequently, and the error rate for such modifiers was a very low 8%. Neither of these effects is due exclusively to rank.

Included in Fig. 3 are also the results from the restricted reading time conditions. Recall was much poorer under these circumstances, but the same dependence upon rank was observed as with free reading.

Sentence Set B. In Fig. 4 the mean reading times for each of the 20 sentences of Set B are shown as a function of the number of propositions

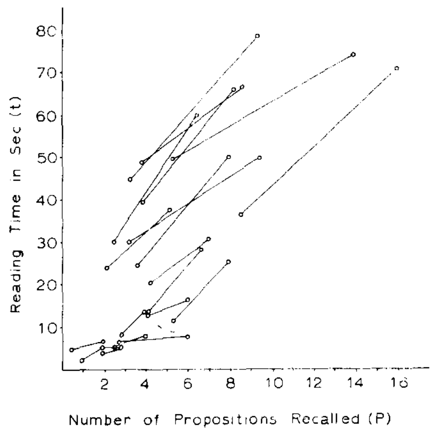


FIG. 4. Mean reading times for the 20 sentences of Set B as a function of the number of propositions recalled. The data for each sentence are shown as Vincent curves.

actually recalled. Figure 4 thus corresponds to Fig. 1, except that each curve is based upon only one-tenth as much data. Because the sentences of Set B differed in length, the data from different sentences could not be averaged as in Set A. In order to obtain stable data, the functions shown in Fig. 4 have been Vincentized. For each sentence the average reading time was determined separately for those trials on which few propositions were recalled (lower 50%, or as close to that as was possible) and for those trials on which many (upper 50%) propositions were recalled. For 19 of the 20 sentences the reading time is longer when more propositions have been recalled than when few propositions have been recalled. Since each curve is based on only one sentence, the number of words as well as syntax are fixed; what differs is the amount of processing done by a subject, as indexed by his ability to reproduce the content of the sentence immediately after reading it. In almost every case more recall meant longer reading times. There is also an indication that as the sentences and paragraphs became longer, the slope of the functions relating the reading time and the number of propositions recalled increases. For the 16-word sentences of Set A, the subjects needed about 1.5 sec per proposition. For the shortest sentences in Set B, those 10 words or less, 1.3 sec were required per proposition recalled. However, for the longest paragraphs (43–58 words) the subjects needed a 4.3 sec average of extra reading time to process a proposition.

The relationship between the number of propositions upon which a sentence or paragraph was based and the average number of propositions which the subjects recalled from that sentence is shown in Fig. 5. For the free reading condition, the results look much like those obtained

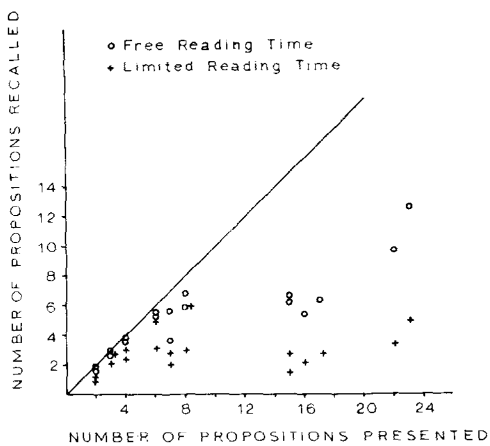


FIG. 5. Mean number of sentences presented and recalled for the sentences of Set B. The standard error for free reading is .41 propositions and is .19 propositions for the limited reading time.

for Sentence Set A: there is a strong positive relationship between number of propositions presented and recalled ($r = .86$). The relationship may not be linear, but at least for the range represented here, the subjects are still able to recall more when they are presented with more. When reading time is restricted, however, this is no longer the case. Except perhaps for the first few points, which are close to perfect recall, there is no noticeable relationship between the number of propositions in a sentence and the amount recalled ($r = .23$, which is not significant statistically). This lack of a relationship occurred in spite of the fact that reading time was longer for the longer paragraphs in proportion to the number of words. Obviously, proportionality was not enough: processing rates decrease more rapidly. Such a result is precisely what one would expect from the free reading data shown in Fig. 4. The slopes and intercepts of the functions in Fig. 4 increase quite dramatically as a function of the length of a sentence, so that giving subjects the same amount of time per word irrespective of total length, as was done here, should not be sufficient for effective processing for the longer sentences. A third of a second per word was quite enough for the shortest sentences, but for the longer paragraphs several times as much would have been required per word to achieve an equivalent level of processing.

DISCUSSION

The purpose of this experiment was to determine whether the number of propositions in the semantic base of a sentence affects the time required by subjects to read that sentence and store it in memory. As Figs. 1 and 4 demonstrated, reading time is indeed a monotonically increasing function of the number of propositions being processed. These data provide support for the model of Kintsch (1972), hypothesizing that the content of a text is stored in propositional form. They extend beyond earlier results, which merely demonstrated the existence and psychological importance of semantic base structures, in that they support a particular type of semantic model, namely one based upon a propositional analysis. Thus, the interest in the present study was not in the basic empirical finding, that is, the longer the study time is for a sentence, the better the recall. On the contrary, this relationship may well be assumed to be true quite in general. The fact that it was possible to maintain this relationship, using the particular analysis of both the text and the recall protocols into propositional units as postulated by the model, may be taken as support for the model. Of course, there may be alternative models and there may be alternative ways to analyze these data: we shall not explore these possibilities here—it is enough to present one feasible approach, given the current lack of satisfactory methods to analyze the comprehension and recall of natural language text.

Are the present data informative enough to permit some inferences as to how information is processed and stored during reading? The general problem is much too complicated and will require extensive research; but the data do offer a reasonable way of approaching the problem, and permit the formulation of some simple hypotheses. A rudimentary information processing model can be constructed to account for some aspects of the reading process. In presenting this model we shall limit the discussion to the data from Sentence Set A. These sentences have been selected because the assumption of constant processing rates appears to hold for them, according to Fig. 1.

Let us accept Fig. 1 as the basis for building the model. We can represent Fig. 1 by

$$t_i = a + bP_i, \tag{3}$$

where t_i is the mean reading time for Sentence i, P_i is the number of propositions upon which Sentence i is based, and a and b are constants. Equation (3) suggests that reading time may be decomposed into two factors: b sec for each proposition read, plus a constant of a sec mean duration. As a psychological interpretation of this equation, one can hypothesize the following. The a sec are a time during which an overall evaluation of the sentence is made, part of which certainly includes syntactic parsing. During each of the b sec periods, a proposition is being processed. Note that it is not necessary that all of the a sec for general analysis precede the processing of separate propositions.

Now let us assume that the processing times for each proposition are exponentially distributed with rate $1/b = \lambda$. This assumption is chosen merely because it is the simplest one known for workable reaction time models. The total processing times for Sentence i would then be composed of a constant plus a gamma distribution with parameters λ and P_i. However, since the subjects may, and do, stop reading at any time, the observable reading time distributions are truncated in unknown ways. This problem could be avoided by considering only the reading times for perfectly recalled sentences, but in the present data, the frequency of such events is much too small to permit an analysis of reading time distributions.

A more practical approach may be taken by noting that if a subject reads a sentence based upon P propositions for time $(t - a)$ with rate λ, the likelihood that he will have processed x propositions during this time is a Poisson variable with parameter $\lambda(t - a)$. But by Eq. (3) and the definition of λ,

$$\lambda(t - a) = \lambda(t - t + bP) = \lambda bP = \lambda \frac{1}{\lambda} P = P$$

so that

$$Pr(x;P) = \frac{e^{-P}P^x}{x!} \quad \text{if } x < P, \tag{4a}$$

$$Pr(x = P;P) = \sum_{x=P}^{\infty} \frac{e^{-P}P^x}{x!}. \tag{4b}$$

Hence one can calculate the likelihood of recall of 0, 1, 2, . . . , P_i propositions for each sentence simply by referring to tables of the Poisson distribution. Equation (4) is concerned with the performance of the average subject since the mean reading time t_i has been used in place of a distribution with considerable variance. But in spite of this simplifying step, Eq. (4) can be used to predict mean recall. Of course, this model is seriously incomplete: it treats all propositions alike, and thus predicts that all sentences of equal word length and of equal number of propositions should be, on the average, equally well-remembered. This is obviously false, but within the design of the present experiment such variability must be treated as error, to the detriment of goodness of fit tests.

In actual fact, this means that the model can be tested only on the basis of data averaged over sentences, so that sentence idiosyncrasies may have a chance to cancel each other. The problem of how to average over recall distributions differing in the number of propositions has been solved in Fig. 6 in the usual way, namely by Vincentizing into quartiles. Predicted and observed recall distributions show a reasonable degree of overlap, but the fit is by no means perfect (the Kolmogorov–Smirnov test detects a significant difference at the .01 level). The source of the problem is easy to see: there are not as many instances of perfect (or near perfect) recall as predicted. This is, however, exactly the kind of deviation from predictions which would be obtained on the basis of individual differences in recallability of propositions: since some propositions are difficult to recall, especially in the longer sentences, perfect recall will be rarer than if all propositions were of equal difficulty.

The extension of the model to the restricted reading time condition is straightforward. If a person reads for exactly 5 sec the proability that he will have processed exactly x propositions, given a sentence i which consists of a total of P_i propositions, is

$$Pr_i(x;\tau(5-c)) = \frac{e^{-\tau(5-c)}((5-c)\tau)^x}{x!} \quad \text{if } x < P_i \tag{5a}$$

$$Pr_i(x = P_i;\tau(5-c)) = \sum_{x=P_i}^{\infty} \frac{e^{-\tau(5-c)}((5-c)\tau)^x}{x!}. \tag{5b}$$

Note that no assumptions about parameter constancy during free reading and restricted reading are made and that new parameters c and τ (corresponding to a and λ) are introduced in Eq. (5). To assume parameter constancy in this case would be quite absurd: subjects have only 5 sec to read each sentence, but the intercept in Fig. 1 was about that much, so unless the subjects speed up their processing rates they could not read the sentences at all.

Unlike Eq. (4), where the parameters could be predicted from an independent set of data, c and τ must be estimated from the observed

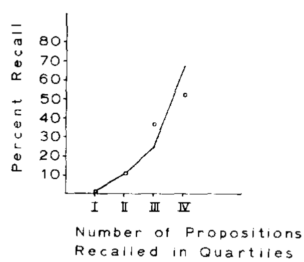


FIG. 6. Predicted (lines) and observed (circles) distributions of the number of propositions recalled for the sentence of Set A, free reading time.

distributions of the number of propositions recalled. Before doing so, however, attention should be called to a very striking feature of Eq. (5): it predicts that the likelihood of recalling x propositions, as long as x is less than the total number of propositions in the sentence, is the same for all sentences, independent of what the total number of propositions actually is. In other words, the probability of correctly recalling 4 propositions from a sentence which contains a total of 5 propositions is the same as the probability of recalling exactly 4 propositions from a sentence which contains 9 propositions. Unfortunately this strong prediction is not testable because of the incompleteness of the model. As already mentioned, the model deals only with the number of propositions upon which a sentence is based, and disregards differences between individual propositions. Hence, if one tests the adequacy of Eq. (5a), deviations from expected values may occur not because the model is wrong, but simply because of its obvious incompleteness.

If the partial recall data from all sentences are pooled, and if Eq. (5a) is fitted to the pooled data by means of a minimum chi-square procedure, the fit obtained is quite good, as is shown in Fig. 7. The minimum chi-square is 14.44, so that for 8 df, $p > .05$. The estimate for the rate parameter $\tau(5 - c)$ is 4.2. If this estimate is used in Eq. (5b), the proportions of complete recall for each sentence can also be predicted. Both predictions and data are also shown in Fig. 7. Obviously, these predictions are poor again. This is because the data are based upon much fewer observations than the partial recall data, and each point was obtained from only one or two sentences, thus again introducing sentence specific effects.

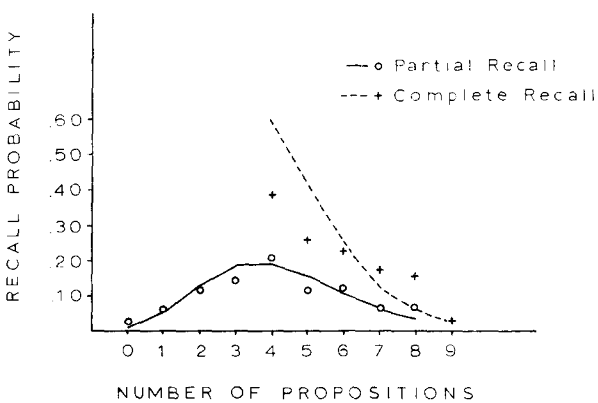


FIG. 7. Predicted (lines) and observed (circles and crosses) recall probabilities for sentence Set A, 5 sec reading time. The partial recall data are averaged over all sentences. The complete recall data are based upon either one or two sentences and have been smoothed by taking running triplet means.

Note, however, that the prediction (and observation) that the likelihood of complete recall decreases as the number of underlying propositions increases in a sentence agrees loosely with results reported by Perfetti (1969). Perfetti found that sentences with a high lexical density, a statistic correlated with though not identical with number of propositions, were recalled more poorly than sentences with low lexical density. There are many other procedural and scoring differences between Perfetti's study and the present experiment, but the consistency of the results should not be overlooked.

From $\tau(5 - c) = 4.2$ one can obtain somewhat speculative estimates of c and τ, if one is willing to assume that the relationship between slope and intercept which was observed in Fig. 1 holds true even if reading time is restricted. Thus, if the intercept is about 4 times the slope, τ becomes 1.64, which compares with $\lambda = .68$ for the free-reading data. In other words, while the subjects took about 1.5 sec per proposition if reading time was unlimited, only .61 sec per proposition was needed when each sentence was exposed for 5 sec.

Clearly, no information processing model will be satisfactory unless differences between propositions are taken into account. However, this limitation, which is inherent not only in the model but in the whole design of the present study, should not detract from the positive accomplishments: number of propositions, although by no means the whole story, has proven to be a useful independent variable for the analysis of both reading time and recall data. One of the next steps will be to explicitly account for at least some differences among propositions. Indeed, although this experiment was not designed for this purpose, some of the subsidiary results discussed above have already provided interesting clues about some characteristics of propositions that are probably important. It has been noted that, in general, superordinate propositions were recalled better than subordinate propositions. One way to incorporate this finding into a model would be to describe a noticing or processing order for propositions which depends upon their hierarchical structure. However, rather than elaborate the model at present, it may be a better research strategy to design experimental studies to explore these problems further. There are, of course, other aspects of propositions not related to their hierarchical structure that were found to be important determinants in recall. Propositions appearing in the surface structure as modifiers were very poorly recalled, while propositions which contained a proper name were extremely well-recalled. In general, it must be remembered that although number of propositions had a large effect upon reading times, it could account for only 21% of their variance. That leaves quite a bit of room for other factors, associated with properties of individual propositions as well as syntactic factors, to operate.

After both free and paced reading, superordinate propositions were recalled better than propositions that were low in the hierarchies that constitute the sentence structures (Fig. 3). This finding is reminiscent of an observation reported by Johnson (1970) that sentence constituents (defined objectively but atheoretically) that were rated as important for the structure of the whole text were also recalled best. However, a theoretically motivated analysis has replaced in the present experiment the judgment of raters.

The role of the syntactic complexity of the material used in the present study needs to be examined further. It has been shown in an earlier study (Kintsch & Monk, 1972) that syntactic complexity may slow down the reading rate, even if propositional content is held constant. One could therefore argue, that the results shown in Fig. 1 may be due, at least in part, to a confounding with syntactic complexity. Increasing the number of propositions in a sentence while holding constant the number of words is necessarily accompanied by an increase in syntactic complexity. Thus, syntactic factors may have been partly responsible for the effects shown in Fig. 1. That such factors cannot wholly, or even largely, be responsible is shown by the data in Fig. 4, where reading time is plotted separately for each sentence. Each curve in Fig. 4 is based upon the same material, so that syntax, as well as the number of words, cannot explain the effects shown.

The data suggest that the processing time per proposition depends upon the length of the text being read. For the briefest sentences used in this experiment, the subjects required little more than 1 sec reading time per proposition, while for 50-word paragraphs about three times as much processing time was required per proposition. It may be that as the length of a text increases, each proposition must be related to more and more others in the text, which could produce such an effect. Or perhaps the longer paragraphs are more complex syntactically than the shorter ones, since no effort to control this factor was made with the present material. Furthermore, interpretations of the increase in reading difficulty with the length of the text are complicated by the possibility that, for the longer paragraphs, some material that had actually been processed might have been forgotten before it could be written down.

The relationship between the present results and the large literature on readability should also be discussed. Chall (1958) had reviewed the somewhat meager results of these studies. Readability, as measured by comprehension tests or simply by subjective judgment, seems to be mainly a question of vocabulary diversity and rarity and sentence length. Attempts to show that "idea density" contributes to reading difficulty have been generally unsuccessful. This has probably been the case because the measures used for "idea density" have been arbitrary and theoretically unfounded (for instance, the most successful index has been the relative number of prepositional phrases in a text). Chall asks "Can one actually dismiss the entire problem of difficulty as one of hard or long words and long sentences?" He clearly does not want to do so, but he lacks a means for quantifying the relevant aspects of a text. More recently, Schlesinger (1968) has shown that sentence length *per se* is not an important variable if some concomitant factors, such as sentence structure and number of words, are controlled. His attempts to show that syntactic factors are important determinants of readability, however, have been largely negative, and he concludes that complexity is primarily a function of content. If one is willing to accept reading rate as an index of readability, the present results hold some promise that more appropriate quantitative measures of content difficulty can be found. The number of propositions in a text seems to be one such measure though, as has been mentioned above, the particular type of proposition as well as structural relations among propositions also play important roles which future research will have to specify.

REFERENCES

BRANSFORD, J. J., & FRANKS, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, **2**, 331–350.

CHALL, J. S. *Readability: An appraisal of research and application*. Columbus, Ohio: Ohio State University, 1958.

JOHNSON, R. E. Recall of prose as a function of the structural importance of linguistic units. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 12–20.

KINTSCH, W. Notes on the structure of semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972. Pp. 247–308.

KINTSCH, W., & MONK, D. Storage of complex information in memory: Some implications of the speed with which inferences can be made. *Journal of Experimental Psychology*, 1972, **94**, 25–32.

PERFETTI, C. A. Lexical density and phrase structure depth as variables in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 719–724.

SACHS, J. D. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, **2**, 437–442.

SCHLESINGER, I. M. *Sentence structure and the reading process*. The Hague: Mouton, 1968.

(Accepted March 13, 1973)