Text Structure and Inference

Walter Kintsch

Department of Psychology

University of Colorado

Technical Report No. 123

Institute of Cognitive Science

University of Colorado

Boulder, Colorado 80309

February, 1983

Text structure and inference

Walter Kintsch

University of Colorado

That inferences have an important function in text comprehension has long been appreciated, but the precise role of inferences in that process is not yet fully understood. Which inferences occur in text comprehension, when, and why? People can infer an almost limitless number of statements from a given text, but they don't produce possible inferences randomly: instead, there are certain statements that almost everyone will infer from a text and others that could be produced with equal justification will never or rarely be generated (Graesser,1981). There are undoubtedly many factors which are responsible for this high level of agreement. The way in which comprehenders organize the text may be one.

The term "inference" is used here in a broad sense. It does not merely refer to statements which can be computed from a text through the application of some rule, e.g. a logical syllogism or pragmatic principles of reasoning, but it comprises instead the whole class of statements which appear in the mental representation of a text but are not directly cued by the linguistic input. That is, it includes lexical inferences as well as schematic inferences. In a schematic inference a schema activated by the text provides default values for components of the schema not covered by the text. The hypothesis explored below maintains that only those defaults play a role in the text representation that correspond to the slots of the schema actually used in forming the textbase.

Other schematic inferences which could be made in comprehending a text will not be made.

This hypothesis derives from the concept of a situational model in van Dijk & Kintsch (1983). We assume that comprehending a text involves the formation of two parallel, complementary structures, the textbase and the situational model. The textbase is a representation of the content of the text, consisting of the elementary propositions derived from the text which are organized into a coherent whole on the basis of some available knowledge structure, e.g. a schema. The situational model is developed in parallel with the textbase proper, but it is a richer structure which contains not only the information derived from the text, but also the reader's previous knowledge about the subject matter. This previous knowledge that is being integrated here with the textual information may be personal or general, e.g. a schema. In the latter case, the schema which is used to organize the textbase provides default values for infromation not specified in the text. Schema based inferences are thus dependent on the overall organization of a text. In forming a textbase higher order propositional units are constructed by assigning the atomic propositions to the corresponding slots of the organizational schema. In general, not all slots of the schema will be filled in this way and those slots of the schema that were not contacted by any textual material contribute their default values to the situational model which thereby remains a coherent structure even though there were gaps in the text itself. Hence, certain components of the schema not mentioned in the text will be inferred in the situational model, while other equally plausible inferences will not be made. That is, it is not sufficient that a schematic inference could be derived from some schema. The schematic inference will only be generated if the corresponding schema is actually controlling the organization of the textbase.

Let A=§a(1),a(2),...a(i),...a(n)† be a script, i.e. a stereotyped action sequence. Given a(i-1) and a(1+1), a(i) can be inferred. Whether it will be inferred depends on whether A is the active knowledge structure which is being used to construct the textbase, i.e. whether the text is "about" A. If not, if a(i-1) and a(i+1) are subordinated under some other structure, say B, a(i) will not be inferred according to the present hypothesis. In terms of a concrete example, consider the routine GETTING A SUNTAN for which norms have been reported by Galambos (1982). Action (6) in that routine is SPREAD OUT THE TOWEL, and Action (8) is WATCH THE TIME; in between is Action (7), LIE ON THE TOWEL. Galambos has empirically determined the components of this sequence, including their order and their centrality. For instance, LIE ON TOWEL has a centrality rating of 7.8 on a scale from 1 to 10, i.e. it is perceived as an important component of the overall routine. Hence, we can be fairly confident that subjects can indeed infer it, given the surrounding statements and the script header. We hypothesize that LIE ON TOWEL will be generated as an inference given GETTING A SUNTAN, SPREAD OUT TOWEL and WATCH THE TIME, if the text as a whole (or rather the relevant text segment) is about someone getting a suntan, but not if it is about something else for which GETTING A SUNTAN merely forms a subordinate episode.

Thus, we would expect LIE ON TOWEL as a schematic inference in the first paragraph in Table 1, which is essentially a paraphrase of nine script actions from Galambos' GETTING A SUNTAN script. According to the model of van Dijk & Kintsch (1983), the GETTING A SUNTAN script will be activated by the first few sentences and phrases in the paragraph, and the whole paragraph is organized by assigning the elementary meaning units derived from the text to the slots of that schema. In parallel with this textbase, a situational model will be constructed in which the components of the schema which did not figure in the

text (e.g. LIE ON TOWEL) are assigned their default values.

Insert Table 1 about here

On the other hand, no such inference should occur in the second paragraph shown in Table 1. The local conditions for the inference are satisfied once again, in that both the script header GETTING A SUNTAN and the two boundary elements PUT DOWN TOWEL and WATCH TIME are present, and hence subjects could produce the inference LIE ON TOWEL. However, the text is not about sunbathing (for present purposes we need not specify what the alternative organization is, as long as it is clearly not the SUNTAN script), and hence the default values of the SUNTAN script are irrelevant.

The experimental method used here to indicate whether or not an inference occurred in a particular context is that of recognition confusions. If subjects tend to falsely recognize a statement expressing a possible inference, we shall take this as evidence that the inference had been produced or would be more likely to be produced in that context. If a potential inference is recognized no more frequently than other distractor items, then we conclude that the inference was not generated during comprehension or was unlikely in that context. Note that the method does not allow us to conclude whether in fact the inference was made during comprehension itself or at some later time up to the presentation of the test sentence.

Method

Subjects. Eighty four introductory psychology students participated in the experiment as part of their course requirement in group sessions.

Materials. The norms of Galambos (1982) were used to construct ten scriptal texts. The norms contain 12 action components for a number of simple scripts or routines, such as GETTING A SUNTAN, CATCHING A PLANE, and PITCHING A TENT. Each action is ranked in the sequence and has a centrality value which indicates its overall importance to the script. From each script, 8-10 actions were selected and simple paragraphs were written based on these script elements, with some minor elaborations to produce a natural if unexciting text. An example is shown in Table 1. In each text, a script element of reasonably high centrality approximately in the middle of the sequence was omitted as a potential inference item. Thus, in Table 1, there are six script elements (CHECK WHEATHER, GET BATHING SUIT, GET LOTION, CHANGE CLOTHES, FIND PLACE, and PUT DOWN TOWEL) before the inferential gap LIE ON TOWEL, followed by three more script actions (WATCH TIME, TURN OVER, RUB LOTION ON). Actions of low centrality and indetermined rank in the sequence were avoided as far as possible. Average centrality ratings for the paragraphs ranged from 5.8 to 7.7.

Each scriptal paragraph was written in two forms which were identical except that the inference item a(i) and the preceding explicit item a(i-1) in one form changed roles in the other form, so that a(i-1) became the inference item and a(i) was explicitly stated in the text. In the example in Table 1 the inference was LIE ON TOWEL and PUT DOWN TOWEL was explicit; in the parallel version LIE ON TOWEL was explicit and PUT DOWN TOWEL had to be inferred. This counterbalancing was done to reduce the effects of ideosyncratic properties of particular inference statements, but will be neglected in the data analysis because the two forms yielded highly comparable results.

For each scriptal paragraph a non-scriptal parallel was written (henceforth, the "framed" and "unframed" versions of the texts). In the unframed versions, the inference item and its two neighbors were taken over from the scriptal texts. In addition, the first sentence or phrase was the same in both versions. The rest of the paragraph contained new material, designed to induce an organization which was different from the simple scriptal structure of the original paragraph. Thus, in the second example in Table 1, the unframed version starts with the same phrase as the framed version, and they both share "She put down her towel and baked in the sun for half an hour", but the unframed text is not based on the SUNTANNING script, as is the case for the framed version. Rather it describes how a girl spent a vacation day. The two versions of each text were of approximately equal length. The ten texts ranged in length from about 100 to 150 words.

For each text pair, eight recognition statements were constructed. There were three old items from the text (always the three sentences or phrases that were common to the two versions, i.e. the first sentence and the two inference inducing statements): "It was such a nice day", "She put down her towel", and "She baked in the sun for half an hour". The correct response on a recognition test for these items was "yes". For the other five test items, the correct response was "no". Of prime interest were the two critical inference items, both paraphrasing the schematic inference LIE ON TOWEL: "She lay down on her towel" and "She stretched out on her towel". Two paraphrases were used to lessen somewhat the likelihood of using an unsuitable verbal expression for the schema slot in question, but also because we wanted to investigate recognition memory for paraphrases of old items without drawing the subjects' attention to these items. Thus, one of the test items was a paraphrase of an old item: "She spread out her towel". The last two test items were a neutral distractor item

and its paraphrase: "Joan wore a colorful summer dress" and "Joan's summer dress was colorful" in the case of the example shown in Table 1. The distractor sentences were plausible or at least conceivable in the context of both the framed and unframed versions of each paragraph (although it was not always possible to make them equally plausible in both contexts), but they were not statements that could be inferred schematically. There were, therefore, four types of test items: "old" (3 for each text), "paraphrase" (1 each), "inference" (2 each) and "neutral"(2 each).

Design and Procedure. There were four groups of 21 subjects, each receiving either the framed or the unframed versions in one of the two parallel forms.

After an unrelated warm-up task, each subject read five of the experimental paragraphs with vague instructions ("we shall ask questions later") and was then given the 40 recognition test sentences from all five paragraphs in random order with instructions to respond "yes" to the sentences that had been seen before and "no" otherwise. After that, the other five texts were read and tested in the same way. Reading time per paragraph was held constant at about 200 words per minute.

Results

Table 2 shows the percent "yes" responses given to the different types of test items after reading the framed and unframed paragraphs. Since the number of items of each type was not constant, the data shown in Table 2 are averages. Thus, the three different old test items are averaged together, as are the two inference and neutral items (in each case, these are paraphrases of each other); there is only one paraphrase of an old item for each text.

Insert Table 2 about here

The likelihood of saying that a text sentence had been seen before differed markedly for the four different item types, ranging from 78% for sentences that had actually been in the text to 8% for the neutral distractor sentences. Paraphrases of old sentences and inferences were in between: subjects clearly distinguished paraphrases (37%) from sentences that they had actually seen, but at the same time they made many more false alarm responses to paraphrases than to the neutral distractor items. For these three types of items, it mattered very little whether subjects had read the framed or the unframed versions of the texts.

A different pattern of results was obtained for inferences: the false alarm rate for inference statements was also intermediate between the false alarm rate for neutral distractors and the hit rate for old items, but it was twice as high when subjects had read the framed paragraphs than when they had read the unframed texts.

A 2x4 between subjects analysis of variance with the factors version (framed and unframed) and item type (old, paraphrase, inference, and neutral) confirmed that this pattern of results is statistically reliable. At a 5% significance level, item types differed significantly, $F(3,162)=336.21$, as did versions, $F(1,162)=8.32$. More importantly, the type x version interaction was significant also, $F(3,162)=10.18$, $MS(error)=4.74$ in each case. A Tukey's test yielded three equivalence groups of means, each differing from all the others but with no significant differences within each group: the two old items formed the first group, the two paraphrase items plus the inference/framed item formed an intermediate group, and the two neutral items plus inference/unframed formed

the third group.

If the test items within each type are not averaged but are considered separately, the results remain essentially unchanged.

Discussion

There are two important results in this experiment: the main result, upon which we shall focus below, is that performance on the inference items was greatly affected by whether subjets had read the framed or the unframed version, and an ancillary result, namely, that subjects distinguished quite well between sentences they had actually seen and their paraphrases. The latter observation agrees with a number of recent studies in which substantial verbatim memory has been found for textual material (e.g. Bates, Kintsch, Fletcher, & Giuliani, 1980). In the present study verbatim memory was found under typical laboratory conditions in which pragmatic factors were probably minimized. Although the retention interval was relatively brief, subjects had read several different texts before being tested for recognition so that short-term memory could not have played a role at all in their performance. It is important to note that subjects showed both memory for meaning ("yes" reponses to the paraphrases of old items were substantially above the false alarm rate) and memory for the surface form of the test sentence ("yes" responses to old items were much more likely than "yes" reponses to the paraphrases of old items). Thus, a model of text memory must account for both meaning and surface effects.

Recognition memory for old statements, paraphrases of old statements, and neutral distractor items was about the same in this study whether these statements were embedded in a scriptal text or not. In contrast, inference items were falsely recognized much more often in the framed texts than in the unframed texts, although the pragmatic conditions for making these inferences

were identical in the two cases. In each cases, a missing step in a script sequence had to be inferred, given the script header and the two surrounding script activities. What was different, however, was the discourse context. In the framed texts the script formed the organizational principle, and hence the default values of the script became part of the textstructure. The inferences, therefore, were frequently confused with statements that were actually present in the text. In fact, inferences elicited about as many "yes" responses as paraphrases of items from the text. The text structure was different for the unframed paragraphs. The scriptal default values played no role here, because the script itself was never activated. Instead, the relevant statements in the text were subordinated under some other organizational schema, with its own defaults and inferences.

According to van Dijk & Kintsch (1983) this result follows from the fact that gaps in the situational model are filled with default values from the schema upon which the text organization is based during comprehension. The present experimental results confirm a central aspect of this prediction, namely, that inferences are context dependent and are controlled by the overall organization of the text. However, they do not address other aspects of the model: the data are neutral with respect to the issue of when such inferences are made (during comprehension or at some later point), and they do not help us to distinguish between the textbase proper and the situational model. Other arguments and experimental results, reviewed in van Dijk & Kintsch, are needed to justify these features of the theory.

At first glance, there appear to be simpler, more parsimonious explanations for the present results in terms of the similarity relations between study paragraphs and test sentences in the two experimental conditions. Inference items at test are similar to all sentences in the original text in the scriptal

paragraphs, but only to two sentences in the unframed texts. Hence there will be more confusion in the former context than in the latter. A number of recognition models would make such a prediction. Consider, for instance, a signal detection model in which the d' values of the test sentences are composed of three additive components: surface memory, memory for meaning, and frame memory. For the scriptal texts, all three factors would be involved in the hit rate for old items, while the false alarm rate for paraphrases and inferences would be derived from meaning-plus-frame memory in the first case and frame memory alone in the second. For the non-scriptal texts, the d' values would lack the frame memory component. Thus, such a model would correctly predict the higher incidence of recognition confusions on inference items in the framed conditions, but at the same time, it would predict a similar difference between framed and unframed conditons for old items, paraphrases, and even for the neutral distractor items. The significant interaction effect observed in the present study clearly contradicts such a model. Indeed, these results rule out all recognition models in which the surface, meaning and frame components are additive, such as a recent recognition model proposed by Gillund & Shiffrin (1982). Their model is developed for list learning data, but the exact form in which it is extended to prose learning is irrelevant for the present argument, because the prediction of a constant difference between all items in the framed and unframed condition is independent of the precise formalization of the model's parameters.

The present results require a recognition model in which the various factors which determine recognition memory are non-additive. One such model would be a sequential decision model, in which a decision is first attempted on the basis of surface memory for the test sentences (which could be successful only for old items, neglecting for the moment erroneous decisions); if no

decision could be reached on that basis, memory for meaning would be examined (with potentially positive results for both old items and their paraphrases); if still no decision were reached, whether or not the test sentence fits into the situational model for the text would be determined (here and here only would the difference between the framed and unframed contexts play a role); finally, a decision on the basis of general similarity between test items and the remembered text is made, yielding the baseline false alarm rate observed for the neutral test items. Such a model readily accounts for the observed interaction effect, and even for the small, non-significant advantage of the old items and paraphrases in the framed condition. Alternatively, similar predictions can be obtained from a random walk model in which the various decision sources compete in parallel rather than sequentially: if memory for surface, meaning, frame, and general similarity compete in a horse-race model with substantially higher rates for surface and meaning than for frame and genaral factors, the pattern of results observed in this study can also be obtained. Detailed fits of these models to the present data would, however, not be very illuminating, because the nunmber of data points to be fitted are relatively small compared to the complexity of the models. Nevertheless, it is of some significance for our understanding of recognition processes to rule out models with additive components in favor of interactive ones.

To explain the present data, therefore, the general framework of the van Dijk & Kintsch (1983) comprehension model is needed (or some theory that makes equivalent predictions in this context), plus some version of an interactive recognition model.

References

Bates, E., Kintsch, W., Fletcher, C.R., & Guiliani, V. The role of pronominalization and ellipsis in text: Some memory experiments. Journal of experimental psychology: Human learning and memory, 1980, 6, 676-691.

Galambos, J.A. Normative studies of six characteristics of our knowledge of common activities. Cognitive Science Technical Report 14, Yale University, 1982.

Gillund, G. & Shiffrin, R.M. A retrieval model for both recognition and recall. Manuscript, Indiana University, 1982.

Graesser, A.C. Prose comprehension beyond the word. New York: Springer, 1981.

van Dijk, T.A. & Kintsch, W. Strategies of discourse comprehension. New York: Academic Press, 1983.

Footnote

This research was supported by a grant from the National Institute of Mental Health, MH 15872. I thank Sheryl Young and Ugo Racheli for their assistance in collecting the data. This is Technical Report No. 117 of the Institute of Cognitive Science at the University of Colorado.

Table 1

Framed and unframed texts for GETTING A SUNTAN

FRAMED

It was such a nice day, 80 degrees and not a cloud in the sky, that Joan decided to spend some time outside. She got her bathing suit out of the back of her closet and the suntan oil from the top shelf of her cabinet. She put on her suit and went out into the backyard . She put down her towel and baked in the sun for half an hour. She carefully turned herself over every couple of minutes so she would get a nice, even tan, and she did not spare the oil. The warm sun made her lazy and sleepy, and when she finally went back into her room she was drained of all energy.

UNFRAMED

It was such a nice day, the first day of summer vacation, and Joan had a thousand things to do that she never had a chance to do during the school year. First, she slept until ten and ate a huge breakfast. Then she went out into the backyard. She put down her towel and baked in the sun for half an hour. She took a shower, put on shorts and a light blouse and went to see Emily. The two girls did not waste much time at home and soon were found strolling down the mall and looking into this and that store, and chatting with some old acquaintances.

TEST SENTENCES

OLD:

It was such a nice day

She put down her towel

She baked in the sun for half an hour

NEW:

PARAPHRASE OF OLD ITEM:

She spread out her towel

INFERENCE ITEM AND ITS PARAPHRASE:

She lay down on her towel

She stretched out on her towel

NEUTRAL DISTRACTOR AND ITS PARAPHRASE:

Joan wore a colorful summer dress

Joan's summer dress was brightly colored

Table 2

Percent "Yes" responses to four types of test items after reading the framed and unframed paragraphs.

	TYPE OF TEST ITEM:			
	OLD	PARAPHRASE	INFERENCE	NEW
TEXT: FRAMED	79	38	27	9
UNFRAMED	77	36	13	7