DOCUMENT RESUME

ED 173 749 CS 004 755

AUTHOR Kintsch, Walter
TITLE On Comprehension.
PUB DATE Apr 79
NOTE 56p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cognitive Processes; Comprehension; Decoding (Reading); *Models; *Reading Comprehension; *Reading Processes; *Reading Research

ABSTRACT

Attempts have been made to develop a model of the process of reading comprehension as a whole that would indicate under what conditions and for what reasons a long sentence might be more effective than several short ones, when repetition is helpful and when distracting, when content is better left implicit in a text, and how overexplicitness might confuse readers. The model relies primarily on two sets of empirical observations -- the time it takes people to read texts and what they can later recall, and attempts to accurately predict reading time and recall data. The reader's goals are represented as schemata, and statistically significant correlations have been found between the following six predictor variables of the model and reading difficulty: the number of reinstatements, word frequency, proposition density, inferences, number of processing cycles, and the number of different arguments. Of these factors, reinstatements, inferences, and the number of processing cycles are deemed most important and worthy of future research. (Numerous diagrams are included to illustrate processes defined by the model.) (DF)

ED173749

ON COMPREHENSION

Walter Kintsch

University of Colorado

Invited Address

AERA Meetings

San Francisco

April 10, 1979

On Comprehension

There is a lot of interest today in the processes that are involved in comprehension, particularly in reading comprehension. Historically, research on reading has focused on the decoding aspects of the process, and we have seen some real achievements in this area. I have heard some very knowledgeable persons maintain that we now have enough knowledge about the decoding aspects of reading, at least for such practical purposes as the design of programs for beginning reading instruction. None, I think, would dare to make a similar claim for reading comprehension. But the time has come when progress can be made in the study of comprehension processes. Efforts in that direction are currently being undertaken by many groups, and it would be interesting to catalogue the various approaches taken and discuss what they have achieved so far. But instead of such a state of-the-art paper, I shall restrict myself in the most parochial manner to our own work: Jim Miller and I, as well as a number of students at Colorado, and Teun van Dijk at Amsterdam, are trying to work out certain aspects of a psychological model of comprehension processes. I would like to describe our work for you, and point out its implications for a problem that has been of continuing interest to educational researchers, namely readability.

The terms "comprehension" and "readability" are not easily defined. At one point in the history of psychology, psychologists talked about the faculties that the mind possesses, such as the "will", "imagination", "memory" or "moral taste". Some remnants of this view are still with us today. We say "Jenny has an IQ of 112", as if intelligence were an inalienable personal possession. Similarly, we might say about a text that "it has a Flesch score of 56". Readability is considered here a property of the text. A number of reading researchers have objected to this practice and have proclaimed an interactionist view, where readability is considered to be the outcome of a reader-text interaction. Most texts are easy to read for some people but hard for others, or easy to read for some purposes but not for others. Stated in these terms, this interactionist position is no more than a truism. For it to become fruitful, some precise ideas about the nature of the reader-text interaction are required. In other words, we need a theory of comprehension; not just miniature models of certain components of comprehension processes, but a model of how the system as a whole works. It would be nice if we could avoid the obvious risks that are involved in modelling a system of such complexity. But the response of a complex system can not always be predicted from the component responses; in order to intervene successfully, we need some idea about how the system as a whole behaves. In general, a text with long sentences is hard to read, but everyone knows by now that this is not always so, and short sentences don't necessarily make a text easy to comprehend. In general, a certain amount of repetition supports comprehension, but quite absurd texts are created by taking this principle too literally. Roger Shuy (1978) has recently discussed some informative examples of this kind. Let me add another one from our own work. Our model focuses on the inferences that a reader must make in comprehending a text. We have proposed the working hypothesis that the more inferences a text requires, the harder it becomes to read. But, as several authors have noted, it is easy to make a text too explicit (e.g. Meyer, 1975; Shuy, 1978). Indeed, we observed some time ago that stating explicitly in a text what every reader would quite readily infer confuses the reader (Keenan & Kintsch, 1974): the overly explicit phrase is falsely considered by the reader as an indication of some complexity in the text that is not there at all, starting a train of useless and confusing inferences. In such a case, the reader tries to figure out what subtle meaning the author had intended by saying what is obvious anyway - except of course that the author had not intended anything in particular and was merely trying to improve readability.

What we need then, is a complete model that tells us under what conditions and why a long sentence might be better than several short ones; when repetition is helpful and when it is simply distracting; when something is better left implicit in a text, and how over-explicitness confuses the reader.

We have been trying to develop such a model, relying primarily on two sets of empirical observations: the time it takes people to read texts, and what they can recall later. Our plan is to design a model so that it predicts reading time and recall data as well as possible, and then to see what this model implies with respect to other interesting issues.

To model "comprehension" in all its complexity is either impossible or very hard to do. But what one can do is decompose the problem into a set of reasonably independent components that can be studied in isolation, and that can then be combined to evaluate their interactions. Thus, we are treating "comprehension" as a partially decomposable system in the terminology of Simon (1974). I shall explain to you what the components of this system are and how we think they work. I shall be brief and informal: a paper published last fall in the Psychological Review by van Dijk and myself describes our theory more adequately, and I refer you to that work for all the details that I must necessarily neglect in this talk. I shall then present some readability and recall data that test one component of the model. Finally, I shall tell you about the further development of the model that is in progress now, and I shall mention another application of the model to a problem of practical educational interest.

The primary goal of our theory is to account for what people remember when they read a text, and to tell what makes a text easy or hard to read. SLIDE 1 illustrates the situation we are dealing with. There are two givens: the text, and the reader's goals and purposes. Actually, we are not really dealing with the text at all; instead we can simplify our task considerably by accepting as an input to our model a semantic representation of the text rather than the real thing. We hand code the text into a set of propositions - conceptual structures that represent the meaning of the text. This is done in a non-arbitrary, codified but non-algorithmic manner. In a sense, this step is outside the model: we need it, but we don't as yet know how to model it explicitly, so we bypass it. This is a weakness of the theory, but I am confident that in a few years the work on semantic parsers that is being done today at several institutions will be so far advanced that we shall be able to borrow their results for our purposes and to incorporate them into this theory.

Note at the top of the slide the reader's goals, which are represented here in terms of a schema. Without such a schema, to control processing our model couldn't work at all. If we are dealing with an unspecified reader, we have no way of predicting what he will get out of a text. Both the reader and the text need to be specified in a model, otherwise comprehension processes are not sufficiently constrained to be theoretically explainable. Many of you will remember that Ernst Rothkopf made this very same point in his talk at these meetings two years ago.

Between the input propositions and the control schema, there is a big black box shown in my slide. The goal of our model is to account for what happens inside that black box. (SLIDE 2) We hypothesize three levels of processing. The input propositions are arranged in a network called a coherence graph. This is done on the basis of some simple coherence rules: connections are formed whenever two propositions share an argument. That is, coherence is defined here in terms of referential coherence alone. At the same time the coherence graph is formed, the propositions are grouped together whenever they belong to the same fact. We might, for instance, know that the driver of an old VW lost control on the ice and smashed into a parked pink Cadillac - which is a lot of propositions, but they all belong to the same fact: an accident, with various participants, circumstances, and modalities. However, not all of the facts related in a text are relevant, and so we need to distinguish further levels of representation, namely the macrostructure of the text. The macrostructure results from the operation of the macro-operators, which are a set of abstraction, or summarization, rules. Indeed, the macrostructure itself may have several levels, corresponding, say, to a long abstract of the text and to ever more concise summarizations. Eventually, merely a title is left.

You will immediately say "Does the world really have to be so complicated - do we really need all these levels?" In order to predict which portions of the input people recall, we need the lowest level. In order to predict the summaries that people make of a text, we need the hierarchical macrostructure. You can think of the fact level as the lowest level of the macrostructure, and I shall later give you a demonstration of the role that facts play in the theory.

Indeed, instead of making the model simpler, I shall need to make it much more complex - for what I have shown you here on this slide has no chance at all of working. In order to make it work, we need direction from above (SLIDE 3). One can't arrive at a summary merely by working upwards from the data level, because we don't know what in the data is relevant and what is not and what can therefore be deleted or generalized. We need a schema that tells us what is relevant, that sets up expectations, that calls for certain facts, inferring them if they are not directly represented in the input set. I have indicated this in the slide by the red arrows. Technically, such a system is modelled with computer techniques that were first used by the Carnegie group on a speech recognition system called HEARSAY. We are using a derivative of that system called AGE which is being developed by Nii and Feigenbaum at Stanford.

But to show you what I am talking about, we don't need a computer. The part of the model that is fully developed at present and from which the readability predictions are derived is the one that contains the coherence rules shown at the bottom of the slide. It takes as its input the set of propositions, ordered as they appear in the text and constructs from it a coherent network, identifying places where inferences are required to obtain coherence. I shall illustrate this construction of a coherence graph with a simple example. In order not to confuse you with technicalities, I shall surpress the propositions in the example and use English text (SLIDE 4) in which word groups that roughly correspond to underlying propositions are circled. Thus, the first sentence reads "The Swazi tribe was at war with a neighboring tribe because of a dispute over some cattle". This sentence constitutes the input to our first processing cycle (SLIDE 5) where a coherent network is constructed with the five input propositions as nodes. The "being at war" proposition is selected as the superordinate of the net (let's postpone the question of how - that belongs to another component of the system with which the coherence rules interact), and the other propositions are annexed to it by a very simple rule: connect all propositions which share a common argument. In the slide, since I am not showing propositions, the way this rule operates is not obvious, but it is an objective, algorithmic rule. Once we have the graph, we procede to the next cycle. But now the question arises which portion of the coherence graph just constructed should be retained in short-term memory so that the graph constructed on the next cycle can be integrated with it. It won't do to construct a separate network each time we read a sentence. We need to interrelate the information from all the sentences in a text. Ideally, we would like to keep the whole tree and simply add the new propositions when we go to another sentence. But people's processing capacities are limited: we cannot keep arbitrarily large amounts of information active in short-term memory. Hence, we select a few propositions on each cycle, keep those in a short-term memory buffer, while all the other, non-selected propositions are relegated to inactive status in long-term memory. We need a rule or strategy that tells us which of the propositions are to be kept in the short-term memory buffer. We have described such a strategy in our paper, which, when used here selects the three propositions circled by the broken red line. The principle behind the strategy is to favor propositions high up in the graph (which tend to be important ones), combined with a recency bias. Again, I shall forego the details, wave my hand,

and go to the next slide. (SLIDE 6) We now have three old propositions in short-term memory, plus a new input, which reads in English: "Among the warriors were two unmarried men, Kakra and his younger brother Gum." The task of the comprehender in this model is to connect these six new propositions with the three old ones still in short-term memory. (SLIDE 7) This fails in an interesting way. There are no concepts shared by the propositions in the buffer and the new input to connect the two sets. So the model says: "Well, maybe I kept the wrong propositions in my buffer. I'll go back and look at everything I have in my long-term memory; maybe there is something there that relates to the new input." This we call a reinstatement search, and we hypothesize that it is one of the things that makes people stumble when they read a text. If a text and a reader interact in such a way that the reader must perform frequent reinstatement searches, then we have a text that is hard to read. In our example the reinstatement search fails: there simply are no common concepts in the first two cycles. So our model builds a new graph, this time selecting "were two men" as its starting propositions and annexing the rest via the concept repetition rule. It also may perform an inference at this point to interrelate the two graphs: e.g., that "the warriors belonged to the Swazi tribe", or that "the people who fight in a war are warriors". This is the kind of bridging inference that Clark and his colleagues have studied quite extensively (e.g. Havilland & Clark, 1974).

Again we select three propositions for our short-term memory buffer, and we go to the next sentence (SLIDE 8): "Kakra was killed in a battle". This time there is no problem (SLIDE 9): we still have "Kakra" in STM, and the new sentence repeats that concept, so we can add it to our graph, select a buffer set, and continue reading. But that is enough for this example. I hope that it enabled you to get some intuitive grasp of how this part of the model works.

What does it do for us? Two things, at this point. It permits us to make some readability predictions: if the model has to make lots of reinstatement searches, lots of inferences, then we predict that people will also have a hard time. Of course, how often the model has to backtrack and go rummaging through its long-term memory depends crucially on the amount of short-term memory we give it. In our example I have assumed a buffer size of 3. Obviously, if I had made it 1 or 2, there would have been more reinstatement searches. In general, the bigger short-term memory, the less trouble the model has with a text. Thus, we need to know what the model's parameters are before we can make any specific readability predictions.

This brings me to my second point. By making some simple assumptions about memory, we can use recall data to estimate the model parameters. Fitting the model to recall data has another advantage: it permits us to evaluate how well the model fits with standard statistical procedures. The logic of this argument is shown on my next slide (SLIDE 10). In the first column we have the 13 propositions from the three sentences I have just shown you. The second column shows how many subjects (out of 100) recalled each proposition on an immediate free-recall test. You see, as is typical, that there is quite a bit of variability there. Now consider the model predictions. The assumption we make is that every time a proposition is processed, it is stored in LTM with some probability *p* (which is another parameter of the model). All propositions are processed at least once (we assume a careful reader here, but this is not necessary), some propositions, however, are processed more than once: for instance, Propositions 1, 2, and 3 were processed in the first cycle and then they were processed again in the second cycle because they were held over in STM. Proposition 10, in fact, had two extra chances to be stored in LTM: it was first processed in the second cycle, held over for the third, and then held over once more for the fourth cycle. I have just indicated these extra processing chances by plus signs in column three, but in the actual model these would be a set of stochastic prediction equations. From here on it is easy going: we find the value of *p*, the probability of LTM-storage, in these equations, which generates the "best" predictions, where best means the minimum chi square on a goodness-of-fit test. If we do this, we obtain a minimum chi-square value in our example of 36.35. For 12 df, this is pretty bad, and clearly indicates that the model predicts a pattern of recall that just was not there in the data. But that is no reason to despair - maybe the model was right, and merely our guess about the size of the STM buffer was wrong! So we try the same thing again, deriving predictions as before except that we restrict STM so that it holds only two propositions. The resulting predictions are shown in the last column of the slide, and you see that now the plus signs seem to correspond to the high values in the data column, as they should. This is borne out by the corresponding minimum chi square of 19.34. This is still not perfect, and if you examine this table a little closer, you see that most of the trouble is caused by proposition 13 (warrior was killed "in a battle"), which only 17 people recalled. The model predicts 46, and this discrepancy greatly inflates our minimum chi square. Proposition 13 is essentially redundant: we know that there was a war, and Kakra was killed, so most people simply don't bother to write down in their recall protocol the redundant information that this happened "in a battle". The Kintsch and van Dijk model actually has a component that deals with this kind of recall suppression: there are production rules that prevent redundant statements from being expressed under certain conditions. Thus, the low recall of Proposition 13 is not really an embarrassment to the model, and is simply irrelevant to the evaluation of the coherence component of the model. If we delete this proposition and recompute our chi square, we obtain a value of 2.81 for 11 df, which is a more valid indication of the fact that, except for small random deviations, our model predicts the pattern of recall remarkably well.

We have followed this strategy to test the model in a large experiment in which 20 paragraphs from various sources were used as the material. The paragraphs, each about 80 words long, were selected so that they would span a wide range of readability. In other words, some of them were pretty awful. Reading time and recall data for these texts were obtained from 120 subjects. So far, we have analyzed the data from half of the paragraphs.

The recall data were analyzed just as I have shown you. That is, only reproductive propositional recall was scored - all errors and constructions were ignored (However, immediate recall with such short texts is known to be almost wholly reproductive). The reliability of the proposition scoring was .91 in this experiment.

A question arises, however, when we come to the operational definition of readability. The measure I prefer is an efficiency statistic: reading time per proposition recalled. Intuitively, this measure appears to be more satisfactory than either reading time or amount recalled alone. However, this measure did not correlate in our corpus with ratings of subjective readability (while uncorrected reading times did), which means that people's intuitions differ about what readability really is.

I now have to give you a little bit more technical detail about the model itself, specifically about the model parameters that were estimated. In my example, I have already introduced one of them - the capacity of the short-term memory buffer. In previous work we have estimated this capacity to be 4 propositions; Jim Voss at the University of Pittsburgh has obtained an estimate of two, but his buffer contained, in addition, some macro-propositions, so the two estimates are not incompatible. Another parameter of the model that we need to be concerned with is the maximum input size per cycle i.e., the number of propositions accepted each time: in the present version of the model, input size is determined by the sentence boundaries in the text, except that when a sentence is too long, a cycle is limited to *I* propositions, where *I* is to be estimated from the data. The third and final parameter of the model that we estimate from the data is *p*, the probability that when a proposition is processed it will be stored in LTM and will be recalled subsequently.

Quite crucial to the model is the strategy that selects the propositions to be retained in the short-term memory buffer from one processing cycle to the next. In the Kintsch & van Dijk paper we described such a strategy. In the present work we have basically retained this strategy, but modified it in some ways. Since these details would mean very little to most of this audience I shall omit them here and refer you to a forthcoming publication of this work. The model, with these small changes, is formalized as a computer program, written in LISP, that accepts as input a proposition list derived from a text with indications of sentence boundaries. For fixed values of the short-term memory and maximum input size parameters, the program processes this input list in the manner hypothesized by the model, thereby generating recall as well as readability predictions. The recall predictions thus obtained are then fitted to the actual data by means of a minimum chi-square procedure, which yields the third model parameter - the learning probability *p* -, as well as a measure of goodness of fit.

Overall, for the 10 paragraphs so far analyzed, these goodness of fit measures were quite satisfactory. The average minimum chi square per text was 60, for 24 df. Although that deviation from the data is highly significant, its absolute size rather than its significance level is more important in goodness of fit tests, and that is fairly good here. Our best-fitted text yielded a non-significant chi square of 28, while most of the texts were in the 40-60 range. Only one paragraph was fitted really poorly, with a minimum chi square of 122. It was the most difficult text of the set, and the model - with only the coherence mechanism at its disposal - simply couldn't handle that one. There was indeed, a general relationship between the goodness-of-fit values and the difficulty of the texts: the easier the text, the better the fit of the model. When you have to start making a lot of inferences, then the coherence rules alone are simply insufficient and need help and guidance from other aspects of the model to which I shall turn presently.

The best estimates of short-term memory capacity were in the range of 3 to 5 propositions, with a mean of 3.88. Maximum input sizes ranged from 5 to 8 propositions, with a mean of 6.2. The mean learning probability was rather high for these short texts, p=.64.

But the most interesting results of this study are the readability predictions. The small variations in the estimates of the short-term memory capacity for the ten different texts appear to be unrelated to readability. The same is true for the learning probability p, which appears to be mostly determined by the proposition density of the text. The third parameter of the model is more interesting: input size correlates .67 with readability, as shown in SLIDE 11. For this slide I have divided the ten texts into a group of 3 easy, 4 medium, and 3 hard texts, based upon their reading times per proposition recalled. For the easy texts more propositions are accepted per cycle than for the hard texts. Input size is also a factor in how long subjects read a text, as would be expected.

The most important factor in the model related to readability is, however, the number of reinstatement searches that are made in processing each text. Reinstatement searches are instances of backtracking: the present input is not related to the propositions that were retained in the short-term memory buffer, the system has to go back and search its long-term memory for a possible relationship that would establish the coherence of the text. SLIDE 12 shows that our three easy texts do not require any reinstatement searches, the three medium texts require on the average .5 reinstatements, but this number increases to 1.67 for the three hard texts. Overall, reinstatements correlate -.62 with readability and about equally with total recall (-.59).

On the other hand, the number of inferences that have to be made in reading a text in order to make it coherent correlate much less highly with the readability statistics. Bridging inferences occur when a concept is not repeated from cycle to cycle, as, for instance, in the example I gave you before where "war" was mentioned in the first sentence, and "warriors" in the second. We know from the work of Clark and others that such inferences require extra processing time, but apparently this effect is relatively weak compared with the extra effort involved in the reinstatement searches. There is an important qualification here: many of the bridging inferences in our material were very easy and obvious, as in the example that I just mentioned. If we arrange things so that the reader does not have the necessary knowledge base to make the inferences that the text requires, they could be just as resource consuming as reinstatement searches, and even more so. Such a case, in fact, occurs in one of our texts where you have to know that Sloan was one of the witnesses who appeared before Judge Sirica in the Watergate trials - information which does not appear to be current among Colorado undergraduates.

I don't want to suggest a new readability formula, but SLIDE 13 is an easy way ~~to summarize~~ our results: in our limited set of data the multiple correlation between six predictor variables and reading difficulty (defined here as the number of seconds of reading time per proposition recalled on an immediate test) is a proud .97. Most of the variance is accounted for by the first two factors - the number of reinstatements, as just discussed, and the traditional word frequency. Two other factors make smaller contributions: proposition density (that is the number of words in the text divided by the number of propositions in its base) and inferences. The number of processing cycles (which is obviously related to the traditional sentence length variable as well as to the maximum input size in the present model) and the number of different arguments make negligible contributions. That word frequency and sentence length are related to reading difficulty is no news. That proposition density and number of different arguments are so related replicates some of our earlier work. About reinstatement searches, inferences, and number of processing cycles we have learned from the present model. I think they are important and interesting factors that deserve a lot more scrutiny in the future.

The Flesch formula by the way, does not predict the readability, as defined here, for this particular set of paragraphs. It does predict raw reading times, however, and is also related to subjective ratings of readability, just to complicate the picture a bit further.

Let me recapitulate what I have talked about so far. I have first outlined a very complex model with several interacting components. I then discussed in some detail one of these components - the coherence rules. I have pointed out the implications that this part of the model has for recall data as well as for readability, and I described an experiment that tested these predictions. Things seemed to be going quite well with this rather simple component of the model, so why then do we need a more complex model, with facts, macrostructures, and all that?

The basic problem is that at the level we have operated on so far our model is simply too stupid. Let's go back to that example of the tribesmen Kakra and Gum; in the last sentence, Kakra was killed. Slide 14 shows the three propositions in the STM buffer after that sentence, as well as a new input sentence:

"According to tribal custom, he was married subsequently to the woman Ami". (SLIDE 15) Everything is fine with the model: the "Kakra" in the buffer reappears in the input sentence, and a nice, coherent graph structure is obtained. The computer doesn't scream, the model doesn't blink an eyelid when a dead man is being married, while you as readers at this point have experienced considerable puzzlement with the text, I hope. That's what I mean with the model being stupid - it is only concerned with the coherence of the text, and doesn't care at all what you say. A real reader will reject this text as nonsensical, or, if he is very imaginative, invent some appropriate tribal custom, or, if he knows about ghost marriages, recognize Kakra's marriage as an instance of ghost-marriage. In fact, my example comes from a study done at Colorado by Donna Caccamise, who was concerned with the effect of knowledge on understanding. She gave her subjects anthropology lessons, in which they learned about the custom of ghost marriages. In a ghost-marriage the oldest son of a family, if he dies without an heir, is legally married, but his younger brother takes his place until an offspring is produced.

Comprehension requires knowledge, and the way knowledge enters into the present model is via the grouping of propositions on the basis of the facts to which they belong. A fact, once established, generates expectations about other related information in the text, as well as about other facts. The way a fact is established in the model is that an appropriate knowledge structure is pulled out of the reader's knowledge store, and the text propositions are related to that knowledge structure. Let us return to my earlier example.

(SLIDE 16): in the first sentence we were told that the Swazi tribe was at war with a neighboring tribe. This activates the "War" frame - the Swazi tribe is inserted as the actor in this war, the neighboring tribe becomes the opponent, there is also a cause. Many other things that we expect when we hear about a war are at this point unspecified, but the system is ready for other information related to this war. Therefore, the next sentence comes as no surprise (SLIDE 17): what we have here is a further specification of the actor in the War - fact. (SLIDE 18) The third sentence adds an outcome. Note that what I am showing here are exactly the same propositions as in the coherence analysis, the only difference is that we are now grouping these propositions in terms of their fact relations. In Cycle 4 finally something interesting happens (SLIDE 19): we are no longer talking about war but about marriage. So the "Marriage" frame becomes the basis for an organization of the input propositions, but it fails: a dead Kakra does not make a good husband. Therefore - if the system knows about ghost-marriage - that frame is called up, and now there are no difficulties, all of the information in the text fits into the slots of the "Ghost-marriage" frame, and no contraditions arise.

What have we gained by this additional analysis? First, consider the readability predictions again. In the coherence analysis we predicted trouble with the transition from the first to the second sentence, because there was no direct conceptual overlap between them. A bridging inference was required - something like "the warriors belonged to the tribe at war". As our data have shown, such bridging inferences do not necessarily make a text harder to read, as long as the reader has available the background knowledge to derive the inferences. My re-analysis of this text in terms of its underlying fact structure gives us a somewhat different picture: the bridging inference is still there, but since the appropriate knowledge structure, namely the war-frame is already activated, it should be an easy inference. On the other hand, a major problem arises when the text shifts from war to marriage: realizing that we are talking now about ghost-marriage is no trivial feat, and readers as well as the model are very likely to experience comprehension problems here. A second advantage of our re-analysis is that the facts provide a basis for dealing with inferences. Although the inference mechanisms themselves are not the focus of our work at present, clearly they will be a very important part of a complete comprehension model. Finally, the fact analysis permits us an easy transition to the topic of macrostructures.

Macrostructures, in this theory, are generated from a text by the reader and correspond to what one usually calls its gist, as expressed by a summary or abstract. By their very nature, macrostructures are hierarchical, corresponding to more and more concise abstracts. It is useful to think of fact-representation as the lowest level of the macrostructure, containing everything that was in the text. The macro-operators pick out from all these facts only those that are relevant. In order to do this, they must, of course know, what is relevant. This guidance is provided by the control schema.

I can not go into any detail at this point, but as always refer you to other publications, giving you just a brief example to illustrate how the model works. Once more, we go back to Kakra and his ghost-marriage to see how the macrostructure gradually evolves during the processing of the text. (SLIDE 20) In Cycle I, when only the first sentence is processed, the hypothesized macrostructure is a simple generalization of the input, deleting everything that appears to be irrelevant at the time: "Two tribes were at war". In Cycles II and III, another macroproposition is added each time: "Kakra and Gum were warriors", and "Kakra was killed". (SLIDE 21) In Cycle IV, however, a complete reorganization is required. Because of the ghost-marriage, different facts now appear to be relevant. The fact that Kakra was unmarried when he died, and that he had a younger brother now appear as the presuppositions for the ghost-marriage, and a new macrostructure is generated by the model at this point, as shown in the slide. Thus, if we asked subjects to summarize our text after the third sentence, they would write something like I showed you on the last slide. But if we asked them to summarize after they had read the whole story, a rather different summary would be obtained (as on the present slide).

But what has all this to do with education? I think it does in two ways. The first I have already indicated in the beginning of this talk. It is the importance of having available a general framework for the understanding of comprehension processes. Educators deal with problems involving comprehension on a day to day basis; they often have to make complex decisions on the spot. Their best guide has been common sense: we are all experienced comprehenders; we all have had our comprehension problems and have developed reasonably good intuitions about them. But common sense will help us only so far. A scientific, theoretical understanding of comprehension processes could be of great help, not in the sense that the theory could solve every one of the educator's problems in real time, but in the sense that it would provide the practician with another set of intuitions about comprehension processes that would sharpen his or her perception and make them more efficient problem solvers.

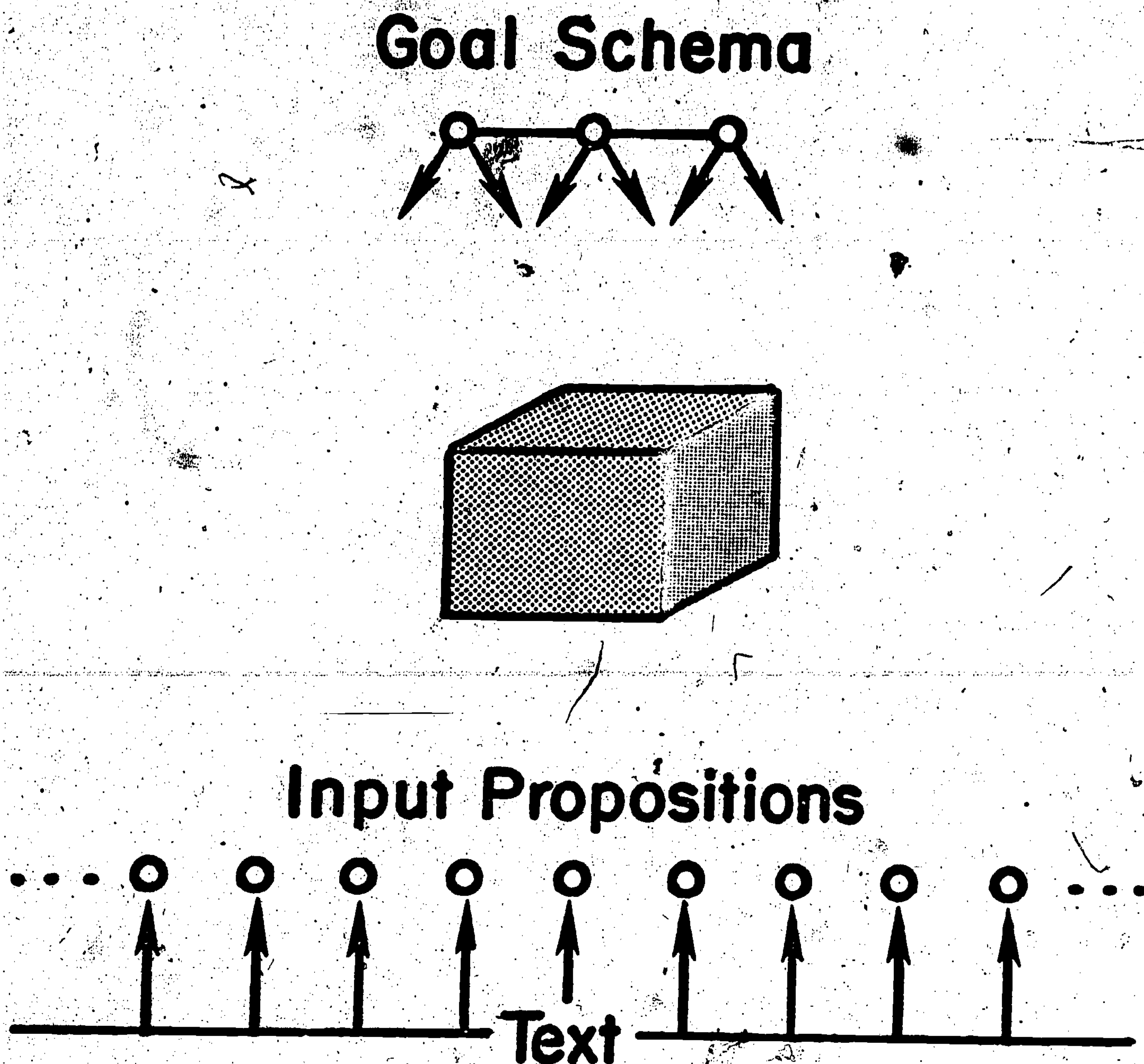
In addition of course, there are the specific results that work like this one produces. I have stressed here the implications of the model with respect to the concept of readability. The model gives us a more refined notion of what is involved in readability, which may eventually have some practical consequences. But that was just one example. It seems to my admittedly biased eyes that wherever I look in the educational literature I find another potential area where this theory could be applied. And that is not surprising. A really general theory of comprehension ought to have strong consequences for numerous educational questions!

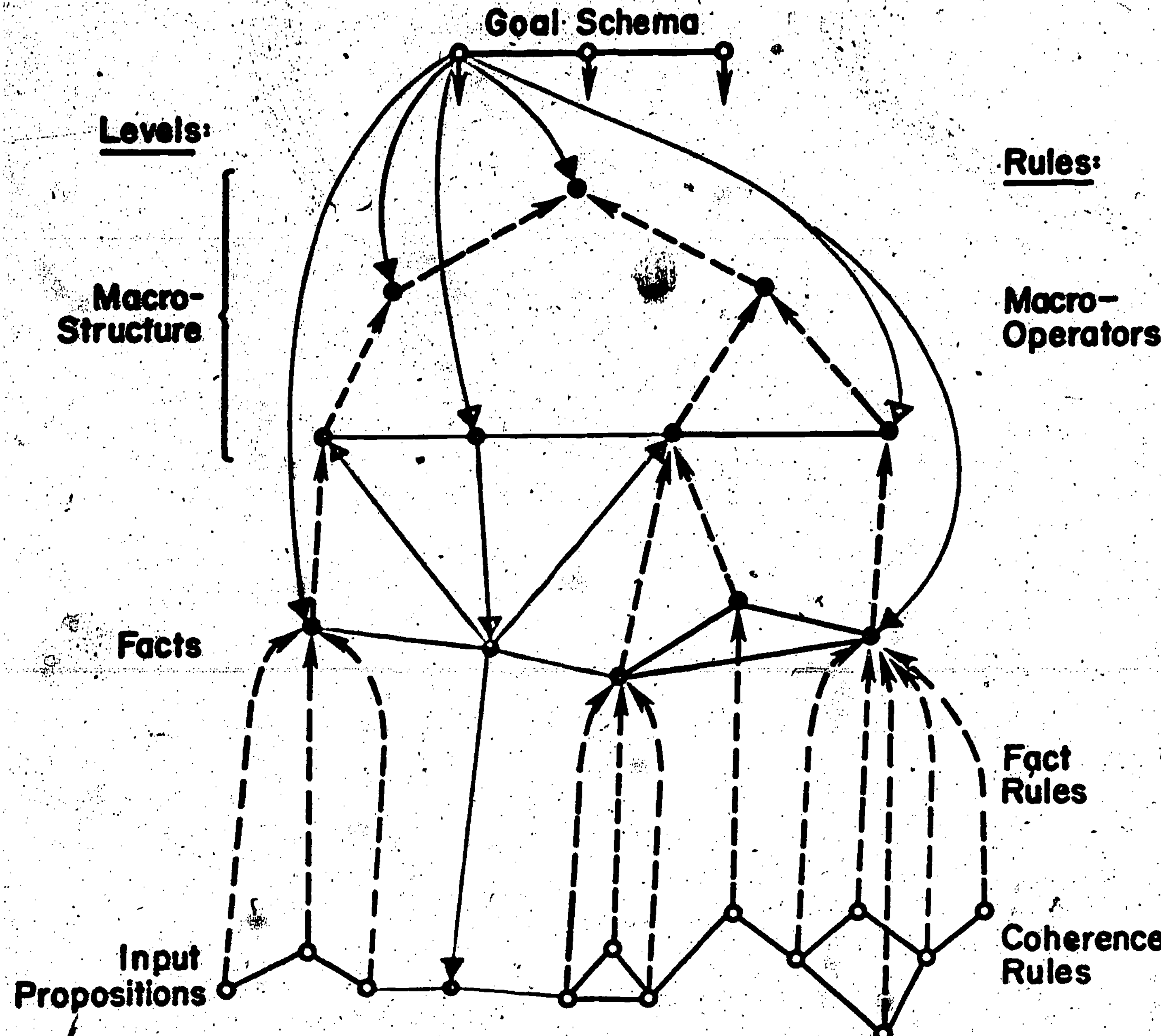
I shall take just a few minutes to tell you about one of these. Mathemagenic behaviors, in particular, asking people questions about what they are reading, have been investigated extensively by educational psychologists. They seemed to hold a lot of promise for improving learning from text, but when all the research was in the results were a little bit disappointing. It was hard to get really spectacular effects, and there were lots of inconsistencies in the data. Now from my point of view, and the advantage of hindsight, this is not surprising if you look at the nature of that research effort. A good review of that work was written in 1975 by Anderson and Biddle. Their main section contained a discussion of the following factors that determine the effectiveness of adjunct questions: nature of test items, positioning and timing of questions, response mode, feedback, overt response, motivation, and finally, just before "other factors" comes the "Nature of the Questions", with only four studies in it, out of a reference list of 3 1/2 pages! But what kind of question you ask ought to be the most important factor of all! The problem is, of course, that in order to investigate this factor one needs to have some kind of theory of comprehension and of the possible roles that adjunct questions might play.

I can not treat this subject here adequately, but I can give you some examples to show you how that could be done. Back, of course, to Kakra and his ghost-marriage. Suppose that after the sentence "Kakra dies in a battle" I had asked the reader "How does the tribe provide for the inheritance of the family property and status when an older son dies without an heir?" In the knowledgeable reader, this question would have primed the ghost-marriage frame and hence insured the correct comprehension of the remainder of the passage. Such a question very well might facilitate retention of new material in the text. On the other hand, suppose we had asked instead, "What was the cause of Kakra's death?" This question does not engage the reader in processing that would help him later on; the answer is merely computed from an already instantiated frame. Therefore, such a question would have very little effect beyond the better retention of the immediately querried item. As a third possibility, imagine a question that would be harmful because it prompts the reader to set up the wrong expectations. For instance, if after the first sentence in our text, which says that there was a war because of a dispute over some cattle, we had asked the reader "What do you think happened to the cattle?", the reader would wrongly focus on the dispute over the cattle, and organize the input propositions in terms of a Dispute-fact, which would have to be replaced in favor of a War-fact in the second processing cycle. Thus, adjunct questions may have positive, negative or null effects, depending upon how the process of question answering meshes with the comprehension process. If I know that at a certain point in the text the reader will need a concept that, according to the model he no longer has available in his working memory, reinstating that concept via a suitable question ought to help. But asking about something that is in working memory anyway will do little good. And if we ask our question really shrewdly, I bet we can confuse the reader, too.

Anderson and Biddle, in the review I just cited, complain that we do not need another demonstration that adjunct questions "work". Instead, we need to know why they work and under what conditions. A model like the present one will let you find out why and when.

Even an incomplete model and one that is undoubtedly wrong in many of its details is better in this respect than no framework at all, because it lets you ask the questions you always wanted to ask and didn't know how.


Goal Schema
Input Propositions
Text


Goal Schema
Levels:
Macro-Structure
Facts
Input Propositions
Rules:
Macro-Operators
Fact Rules
Coherence Rules

Input Sentence 1:

The Swazi tribe

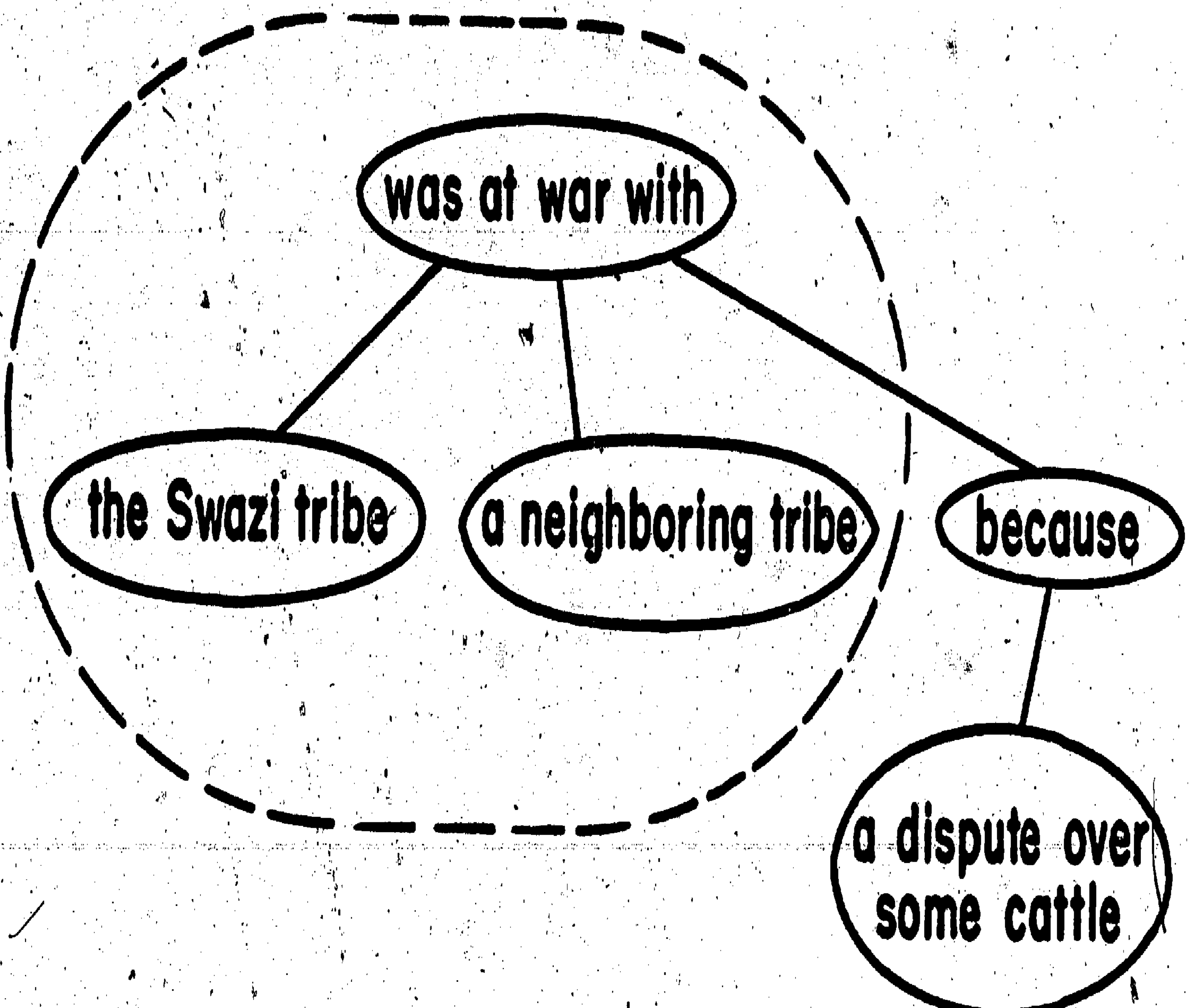
was at war with

a neighboring tribe

because of

a dispute over some cattle.

Coherence Analysis - Cycle I:



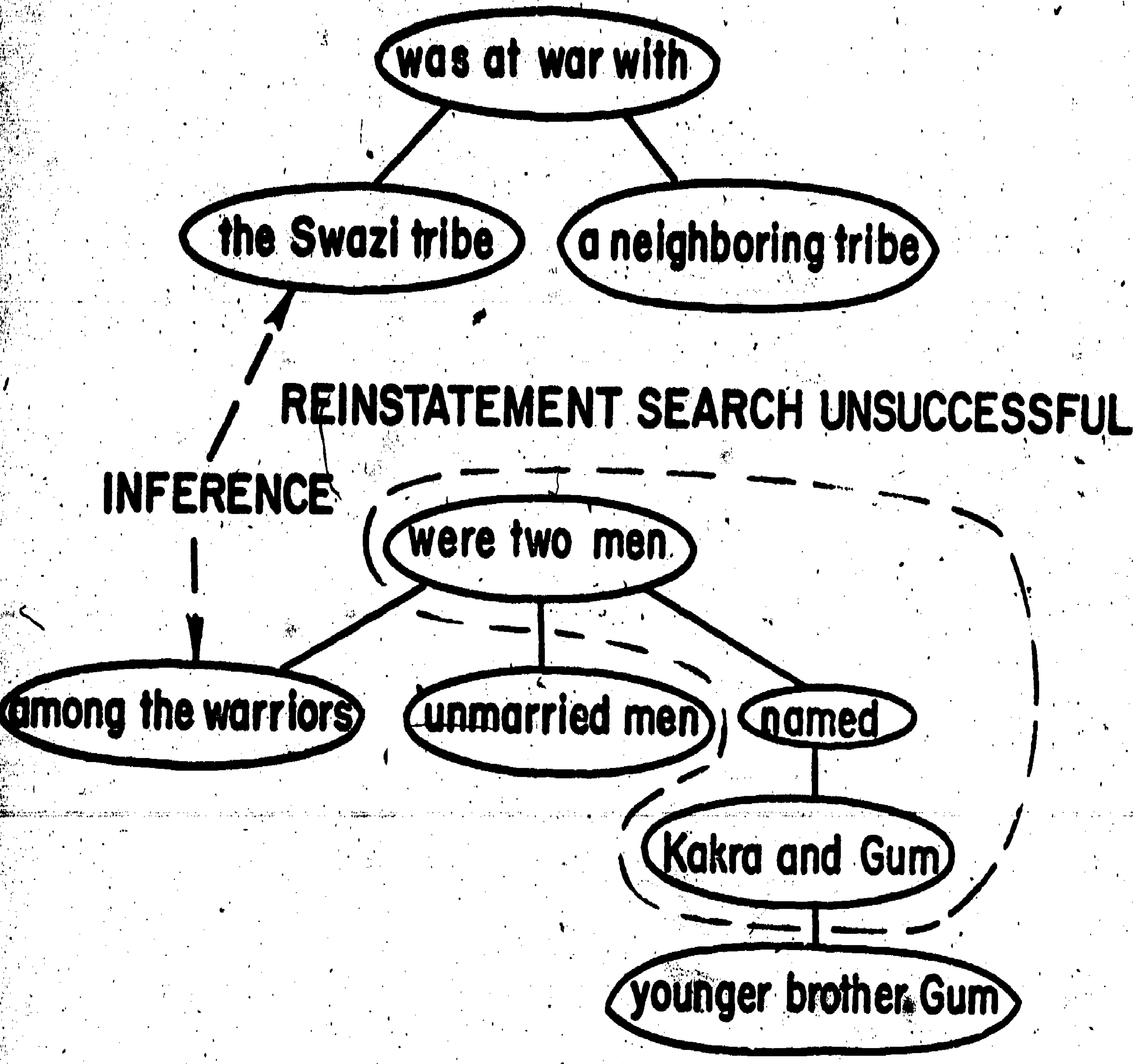
Short-Term Memory:

Input Sentence 2:

Among the warriors | were two men | unmarried men

named | Kakra and Gum | younger brother Gum

Coherence Analysis – Cycle II:



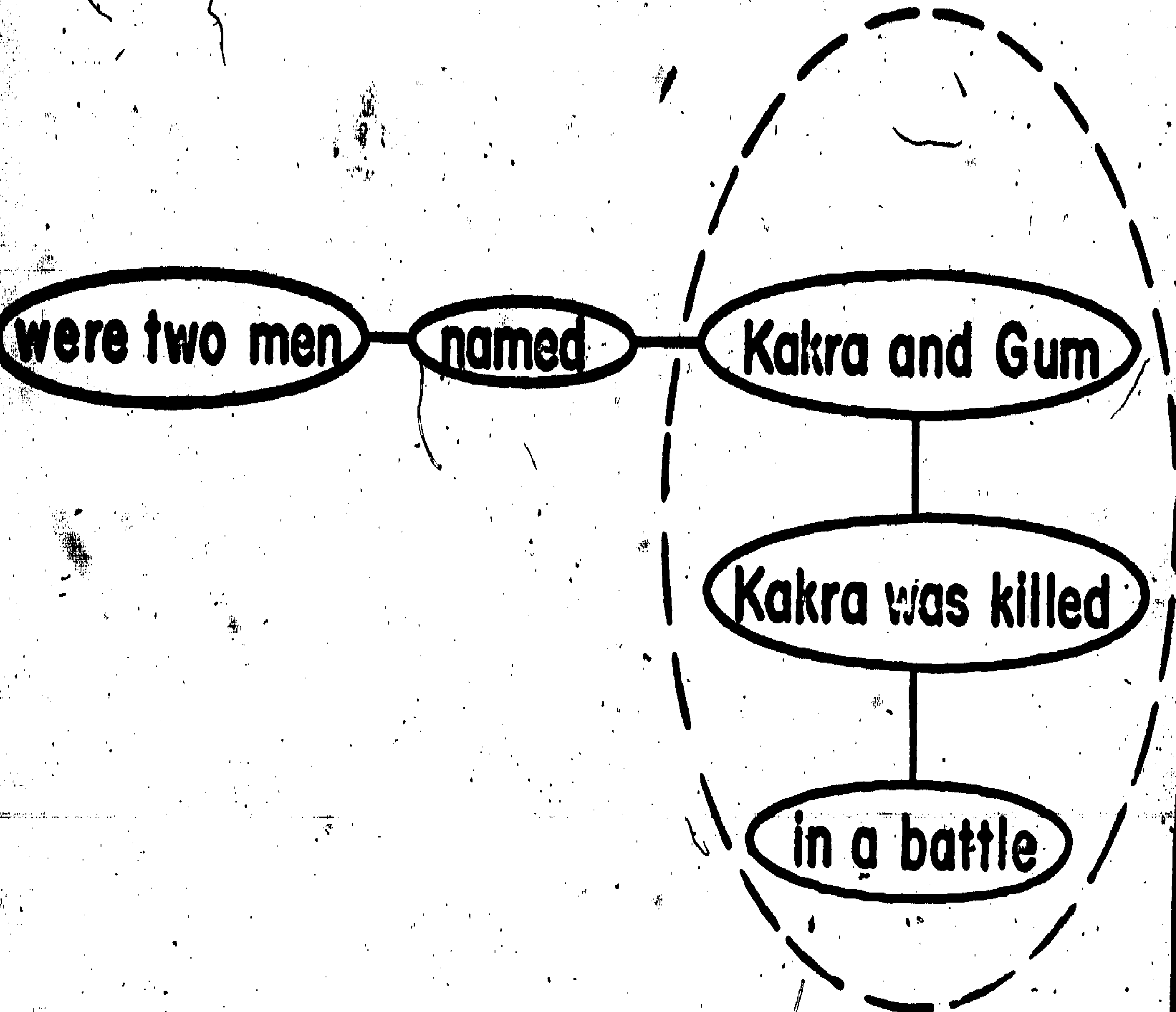
Short-Term Memory:

were two men — named — Kakra and Gum

Input Sentence 3:

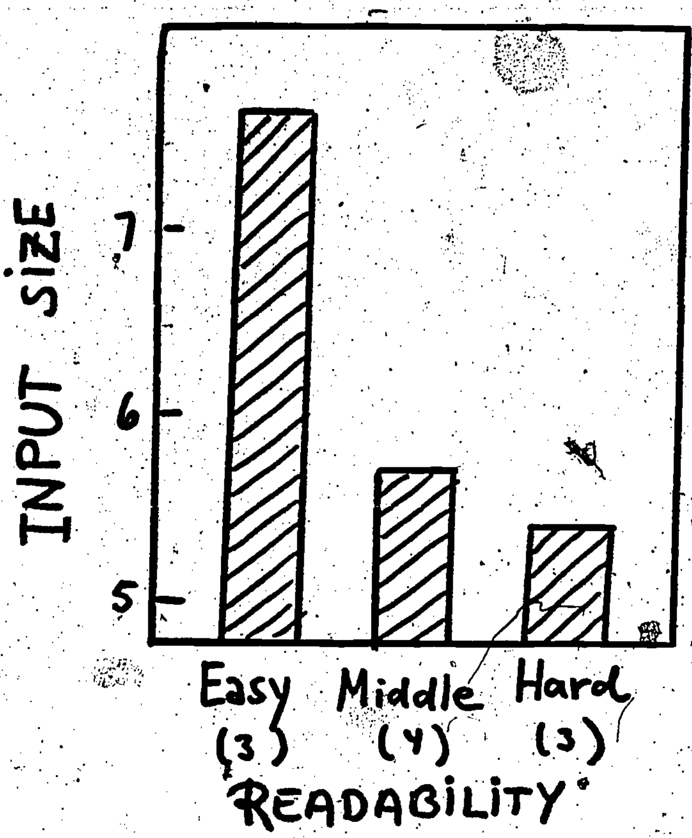
Kakra was killed   in a battle.

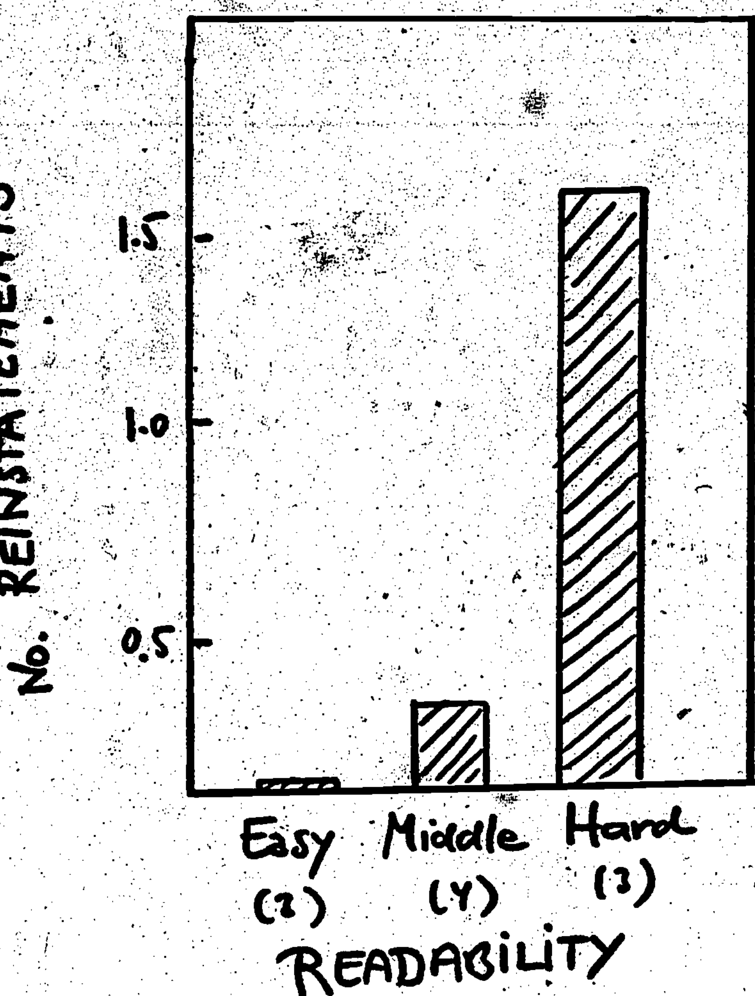
Coherence Analysis - Cycle 3:



Kakra-Text:
Data and Predictions

PROP	RECALL (100 Ss)	PRED S=3	PRED S=2
1	45	+	
2	80	+	+
3	78	+	+
4	46		
5	39		
6	42		
7	82	+	+
8	47		
9	79	+	+
10	81	++	+
11	45		
12	84	+	+
13	17	+	


INPUT SIZE
7
6
5
Easy
(3)
Middle
(4)
Hard
(3)
READABILITY


No. REINSTATEMENTS
1.5
1.0
0.5
Easy
(2)
Middle
(4)
Hard
(3)
READABILITY

READING DIFFICULTY =

$$= 2.83 + .48\,(\text{REINST}) - .69\,(\text{WORDFRQ}) + $$

$$+ .51\,(\text{PROPDENS}) + .23\,(\text{INF}) + $$

$$+ .21\,(\text{CYCL}) - .10\,(\text{ARGUM})$$

Short-Term Memory:

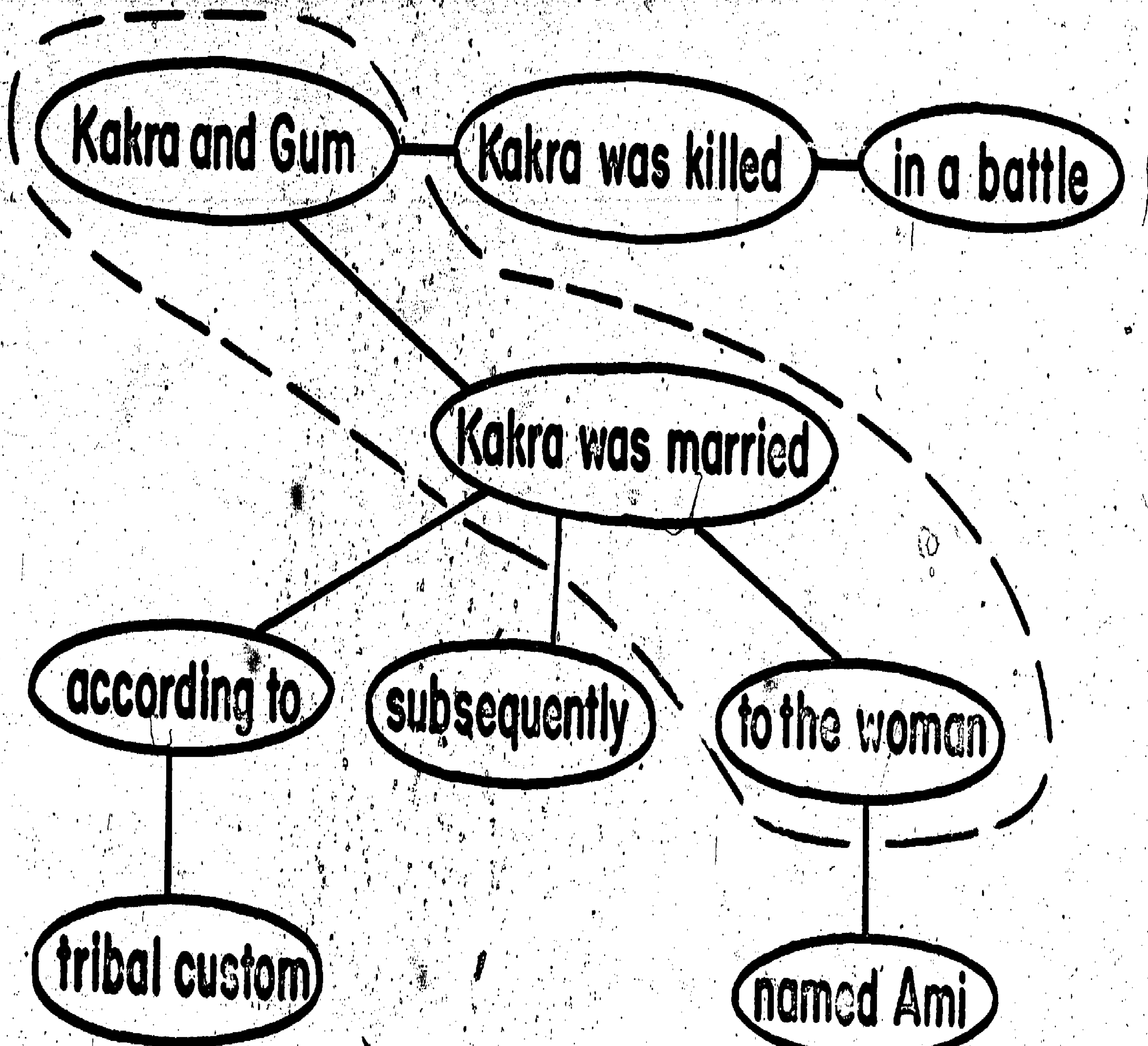
Kakra and Gum — Kakra was killed — in a battle

Input Sentence 4:

According to | tribal custom | he was married

subsequently | to the woman | named Ami.

Coherence Analysis – Cycle 4:



Cycle I:

WAR: (was at war with)

Actor: (the Swazi tribe)

Opponent: (a neighboring tribe)

Cause: (because of)(a dispute over some cattle)

Outcome:

⋮

Cycle II

WAR: (was at war with)

Actor: (among the warriors)
(of (the Swazi tribe))
(were two(unmarried men))
(Kakra and (his younger brother Gum))

Opponent: (a neighboring tribe)

Cause:

Outcome:

•
•
•
•

Cycle III:

WAR: (was at war)

Actor: (two men) (Kakra and Gum)

Opponent:

Cause:

Outcome: (Kakra was killed)

(in a battle)

. . . .

Cycle IV:

MARRIAGE: (was married)

- **Husband:** (Kakra) !!!!(Kakra dead)
- **Wife:** (the woman (Ami))
- **Modality:** (according to)(tribal custom)
- **Time:** (after WAR)
- ⋮

GHOST-MARRIAGE: (was married)

- **Dead Man:** (Kakra)
- **Younger Brother:** (Gum)
- **Wife:** (the woman (Ami))
- **Modality:** (according to (tribal custom))
- **Time:** (after WAR)
- ⋮

Macrostructure:

Cycle I:

(two tribes were at war)

Cycle II:

(two tribes were at war)

(Kakra and Gum were warriors)

Cycle III

(Kakra and Gum were warriors)

(Kakra was killed)

Cycle IV :

(Kakra and his younger brother Gum were unmarried)

(Kakra was killed)

(Kakra ghost-married Ami)