Cognition, 11 (1982) 29–46
© Elsevier Sequoia S.A., Lausanne – Printed in The Netherlands

Referential continuity and the coherence of discourse

ALAN GARNHAM
JANE OAKHILL
P. N. JOHNSON-LAIRD*
University of Sussex

Abstract

Two experiments were carried out to investigate the role of referential continuity in understanding discourse. In experiment 1, a group of university students listened to stories and descriptive passages presented in three different versions: the original passages, versions in which the sentences occurred in a random order, and randomised versions in which referential continuity had been restored primarily by replacing pronouns and other terms with fuller and more appropriate noun phrases. The original stories were remembered better, and rated as more comprehensible, than the random versions, but the restoration of referential continuity ameliorated the effects of randomisation. The descriptive passages had little referential continuity from one sentence to the next, and as expected the effects of randomisation on comprehensibility and memory were negligible. In experiment 2, a group of skilled comprehenders and a group of less skilled comprehenders were selected from a population of 7–8-year-old children. The difference between the groups was known to be largely their inferential ability in reading texts. Both groups read a series of short stories presented in the same three versions as used in the previous experiment. As predicted, the ameliorating effects on memory of restoring referential continuity in a randomised story were confined to the skilled group. The results are discussed in relation to the theories of story grammar, text microstructure, and mental models of discourse.

---

*A number of the arguments against story grammars which we discuss in the introduction are part of an 'oral tradition'. We know that similar points have been made by Emmon Bach and Keith Stenning. Experiment 1 was carried out by the first author while he was supported by the Sloan Foundation at the Center for Cognitive Science at the University of Texas at Austin. Experiment 2 was carried out by the second author while she was supported by the Social Science Research Council at the Laboratory of Experimental Psychology at the University of Sussex. The third author was also supported by the Social Science Research Council. Reprint requests should be sent to P. N. Johnson-Laird, Centre for Research on Perception and Cognition, Laboratory of Experimental Psychology, University of Sussex, Brighton, BN1 9QG, England.

What makes discourse *discourse* as opposed to an incoherent assemblage of sentences? Obviously, there must be constraints on the way in which sentences are put together to form larger units, and equally obviously people must have at least an implicit knowledge of these constraints, as is evinced by their ability to talk and to write coherently, to understand discourse, and even to make judgements about its coherence. The problem, however, is to discover both the nature of the constraints and how they are put to use in the production and interpretation of discourse.

A major proposal in the literature, which has inspired much research, attempts to extend Chomsky's notion of grammar to deal with the structure of stories (see e.g., Mandler and Johnson, 1977; Rumelhart, 1975; Thorndyke, 1977). Chomsky (1957, 1965) proposed that the set of well-formed sentences of any language, together with their syntactic structures, could be generated by a wholly explicit set of recursive re-writing rules. Likewise, a *story grammar* comprises a set of rewrite rules, such as:

STORY ⟶ SETTING THEME PLOT RESOLUTION
THEME ⟶ (EVENT)* GOAL

and, just as a sentence grammar assigns parse trees to sentences, so a story grammar assigns hierarchical tree structures to stories. But, unlike sentence grammars, grammars for stories face four crucial problems.

First, the terminal nodes in the structural tree for a story are assumed to be filled by propositions, but there are indefinitely many propositions which could occur in such categories as OUTCOME or ATTEMPT. There can be no finite listing of such propositions, and no story grammarian has ever provided an explicit set of principles for determining which propositions are members of which classes. It is doubtful whether any such principles will ever be forthcoming. Sentence grammars are indeed very different: their terminal nodes are filled by morphemes drawn from fixed classes, whose members can be specified by enumeration.

Second, there are clearly some constraints which must be placed on relations between the propositions expressed by the terminal nodes in a tree. Some of these constraints might have a 'syntactic' flavour, but most of them concern continuity of content, and seem better classified as 'semantic'. Thorndyke (1977) does not deal with this problem at all, so his grammar would generate a 'story' which comprises the beginning of Snow White, the middle portion of Cinderella, and the end of Sleeping Beauty. Rumelhart (1975) has attempted to place semantic constraints on the relations between the propositions. He proposes rules such as:

EPISODE ⟶ EVENT REACTION
⟹ INITIATE (EVENT, REACTION)

But he does not say how to determine when one event INITIATEs another.

Again, a finite listing of all such pairs of events would be impossible, but, unless explicit criteria can be provided, there is no way of deciding whether a given tree is generated by the story grammar.

Third, the majority of story grammars make use of a *context-free* phrase structure grammar. Grammars of this type are more powerful than finite-state grammars, and less powerful than tranformational grammars. Hence, a case must be made to justify their use rather than grammars of a stronger or weaker generative power. This issue has only recently been recognised by story grammarians, but it is important. Certain sorts of finite-state grammar directly correspond to the Markovian processes favoured by psycholinguists of the pre-Chomskyan era. These processes can be quite adequately carried out by the mechanism embodied in a 'habit-family' hierarchy (see Miller, 1967, who uses Michael Frayn's delightful idea of a machine for writing newspaper stories to illustrate a finite-state grammar).

It is sometimes said that a finite-state grammar for stories would not suffice, because it could not accommodate self-embedding (Black and Wilensky, 1979). This claim must be stated with care: on the one hand, finite-state devices can deal with any *finite* number of self-embeddings, and on the other, a finite-state device that permitted *any* sequence of constituents would plainly accommodate self-embedding structures. A proof of the inadequacy of finite-state grammars for stories depends on showing (i) that an indefinitely large number of self-embeddings can occur within a single story, and (ii) that any such grammar which generated all legal self-embedded stories would also admit some non-stories (Levelt, 1974). The proof is not easy, and it certainly has not been established for story grammars. Likewise, Black and Wilensky's claim that a context-free grammar could not accommodate discontinuous elements is incorrect. If the node labels of a context-free grammar are construed as complex symbols, rather than as unanalysable primitives, discontinuous constituents can be handled quite readily (Stanley Peters, personal communication; Gazdar, in press).

Fourth, story grammars are usually intended to deal with only a restricted set of stories, e.g., genre stories that possess a stereotyped and repetitive structure. But, if there is no independent way to specify the set of relevant stories, the whole exercise becomes dangerously circular: the grammar is intended to analyse just those stories that fit its rules.

Of course, actual story grammars appear very plausible, but this plausibility, we contend, derives largely from what users of the grammar bring to it, rather than from any explicit formulation of principles in the grammar. In many cases it is quite obvious when one event is the OUTCOME of another, but only explicitness about the membership of such classes can give story grammars explanatory power.

The virtue of story grammars is that they attempt to formalise intersentential relations which are used in understanding text. Their major empirical claim is that certain types of story have a specifiable structure, which is independent of their content, and which people know and use in the course of comprehension (Mandler and Johnson, 1980). However, there are many other forms of discourse that are entirely coherent, even though they fall outside the domain of stereotypical stories. Our main aim in the present paper is to investigate what makes texts in general coherent, since the same underlying principles may well do the work supposedly done by story grammars.

There appears to be one overriding and necessary condition for the coherence of discourse: it should be possible to construct a unitary representation that integrates all the information carried in its separate sentences. This condition in turn requires that the sentences make reference, either explicitly or implicitly, to referents in common. Discourse is sometimes about a single major topic to which the majority of sentences make reference, but perhaps more often a series of referents is introduced from one sentence to the next. The ease of establishing what these referents are depends on the referential continuity of the text. Ehrlich and Johnson-Laird (1980) have shown that a referential discontinuity—the occurrence of a sentence expressing a proposition that can only be integrated subsequently—has a disruptive effect on both comprehension and memory. Coherence also depends on abiding by the general co-operative principle that sentences should be structured so that they can be readily understood (see Grice, 1975). For example, information that is taken for granted should either have been previously established or be readily inferable. However, given that there are common referents, and that they can be recovered without difficulty, then how sensible a discourse seems will depend on the plausibility of the actions, states, and relations, in which the referents participate, and on the order in which these events occur. If a text reports an unlikely sequence of singularly implausible events, then strictly speaking it is not incoherent, but rather the world it describes is bizarre. We accordingly distinguish between the coherence of a discourse and its plausibility. Coherence depends on common referents, referential continuity and general adherence to the Gricean co-operative principle. Plausibility depends on verisimilitude to known intentional, causal, and temporal relations between objects in the world.

One of the main pieces of evidence on which psychological claims about story grammars are made is that jumbled versions of stories are more difficult to understand than the original stories. The usual explanation of this finding is that jumbling the sentences in a story breaks up its overall struc-

ture, and hence it cannot be parsed by a story grammar. Thus, one of the reader's main guides to understanding, a knowledge of the rules of story grammar, is no longer of any use in the processing of such passages. However, it is clear that jumbled versions of stories lack both continuity of reference and a plausible sequence of events (insofar as they can be understood in the absence of referential continuity). Thus, for example, a pronoun in a jumbled text may entirely lack any plausible antecedent. There is evidence that it takes time to find an antecedent for an anaphoric expression (e.g., Garrod and Stanford, 1977; Haviland and Clark, 1974); if the task becomes impossible then readers will be unable to form an integrated representation, and their processing of subsequent text may be impaired by a prolonged search for possible antecedents. They may even realise that the story has been scrambled, and cease to treat it as a coherent text.

In Experiment 1, we examined the comprehension and memory of randomised texts in which referential continuity had been restored. A preliminary study had confirmed that randomised stories of this revised variety were better remembered than unmodified randomisations, though not as well remembered as the original stories. Many descriptive passages have a more diffuse coherence than stories: there is little immediate referential continuity from one sentence to the next, but rather each sentence takes up a different aspect of the topic. We can therefore predict that randomisation should have relatively little effect on performance with such texts and, furthermore, that any effects should disappear when referential continuity is restored. Experiment 1 compared such descriptions with stories, and examined the comprehensibility and memorability of three versions of them: the original texts, versions with the sentences in random order, and randomised versions in which referential continuity had been restored primarily by replacing pronouns and other isolated anaphoric expressions with more complete and more appropriate noun phrases.

Experiment 1

Method

Subjects

24 subjects took part in the experiment for course credit. They were undergraduates enrolled in an introductory psychology course at the University of Texas at Austin. All were native speakers of English.

Materials

Six passages, each 200–250 words in length, were used in the experiment. Three were folk tales, of the kind which might be expected to receive a story grammar analysis, and three were descriptive passages. The passages, except for *The Old Farmer and his Stubborn Animals*, were modified to some extent, so that they were all approximately the same length. The sources of the materials were as follows:

(1) The Old Farmer and his Stubborn Animals—version used by Thorndyke (1977).
(2) The Hedgehog and the Jackal—shortened version of that found in Greenway (1965).
(3) The Fox and the Eagle—shortened version of that found in O'Sullivan (1966).
(4) New Year in Hong Kong—slightly modified version of the beginning of Cure (1980).
(5) The Giant Anteater—slightly modified version of the beginning of d'Aulaire and d'Aulaire (1979).
(6) The Statues on Easter Island—slightly modified version of the beginning of the section on 'The Large Stone Statues' (pp. 75–81) in Putigny (1976).

We constructed a second version of each passage in which the sentences occurred in random order but were otherwise unchanged. We then constructed a third version of each passage in which the sentences were left in the same random order as in the second version, but minor changes were made to them. Most of the changes were designed to re-establish *referential continuity* and to ensure that appropriate referents could be found for anaphoric expressions. Three main kinds of alteration can be distinguished.

(a) Where coreferential expressions had become separated, so that they could no longer reasonably be construed as denoting the same object, some clear indication was given that separate individuals or events were intended, for example:

The farmer was *again* trying to put his donkey into its shed.
If you will get me *still* more milk.

(b) Where expressions with antecedents now appeared without them, and where no antecedent could reasonably be inferred from the context, they were changed appropriately, for example:

The farmer's dog refused to bark. (Rather than: But the dog refused.)
When the giant anteater swims. (Rather than: When it swims).

(c) Phrases such as *there was once* which had been moved from their appropriate place in the passage were changed.

The three versions of each passage were recorded on tape by a female native speaker of American English at normal reading speed. Following Thorndyke (1977), a short pause was introduced between each sentence so that the normal versions were not easier simply because reference resolution was aided by the way they were spoken.

Design

Each subject heard six passages, one version of each story and one version of each description. They heard a normal, a random, and a revised random story, and similarly one of each kind of description. Each block of three subjects heard all the combinations of passages and versions. The stories and the descriptions were interspersed with one another, and the first and last three passages which a subject heard comprised one of each of the three versions. Subjects in each block of three heard the passages in reverse order to the corresponding subjects in the previous block, in order to counterbalance for carry-over effects.

In short, passage version was a within-subjects, within-materials variable; order of presentation was a between-subjects, within-materials variable; and the kind of passage (story *versus* description) was a within-subjects, but between-materials variable.

Procedure

The subjects were tested in groups of 1–4 in a small seminar room. They were told that they would hear six passages and that after they had heard each passage they would have to make a judgement about how easy it was to understand, and then to write down as much of the passage as they could remember. The first passage was then played to them, and they were asked to rate its comprehensibility on a scale from 1–10, with 1 meaning 'very hard to understand'. Following Thorndyke (1977), the subjects were asked to take into account both how well the passage formed a coherent whole, and how well the combination of ideas formed a sensible text.

Immediately after the comprehensibility test the recall test was given. Subjects were asked to write down the exact words they had heard, in the order in which they had heard them. However, they were encouraged to use their own words for what they could not remember exactly, and to write down material out of order, if they could not remember where it occurred. A time limit of seven minutes was placed on this task. The same procedure

was followed for the other five passages, and finally the subjects were debriefed. The experiment lasted for about 50 minutes.

Scoring procedure

Previous studies (e.g. Thorndyke, 1977) have scored the number of propositions from the original passage that are present in recall. However, it was difficult to divide up the descriptive passages into propositions in a principled way. Furthermore, the preliminary study, in which only stories were used, established that the number of propositions recalled yielded the same pattern of results as simply counting the number of words in the recall protocols. In the present experiment there was much more variation in the length of the passages. We, therefore, converted the number of words in each recall protocol into a proportion of the number of words in the corresponding passage. This proportion was the main dependent variable.

Results and discussion

Table 1 shows the mean comprehensibility ratings (out of 10) for the normal, random and revised random versions of the stories and descriptions. The version in which a text occurred had a significant effect on its rated comprehensibility (analysis by subjects: $F = 62.06$, d.f. = 2,44, $p \ll 0.001$; analysis by materials: $F = 5.52$, d.f. = 2,4, $p = 0.071$). But there was an interaction between the kind of passage and the version in which it was presented (min$F' = 6.54$, d.f. = 2,7, $p < 0.05$). On the one hand, there was a reliable effect of version on the judgements about the stories (min$F' = 33.74$, d.f. = 2,19, $p < 0.001$): the original stories were rated as more comprehensible than the revised random versions of them (min$F' = 32.38$, d.f. = 1,16, $p < 0.001$), and both these versions were rated as more comprehensible than the simple random versions of the stories (min$F' = 35.42$, d.f. = 1,16, $p < 0.001$). There was also a reliable trend over the three versions (Page's $L = 326.5$, $z = 5.56$, $p \ll 0.001$). On the other hand, there were no reliable effects of version on the rated comprehensibility of the descriptions.

Table 2 presents the mean length (in words) of the recall protocols for the normal, random, and revised random versions of the stories and descriptions. The stories were better recalled than the descriptions (min$F' = 75.14$, d.f. = 1,20, $p \ll 0.001$). The version of a text also had a significant effect on memory (min$F' = 10.72$, d.f. = 1,7, $p < 0.05$); but the two main variables interacted significantly (min$F' = 7.27$, d.f. = 1,5, $p < 0.05$). As with com-

Table 1. *The mean comprehensibility ratings (out of 10) for each sort of passage and version in Experiment 1*
Each entry is summed over 24 subjects and 3 passages.

	Original	Revised	Random	Overall
Story	8.8	4.6	3.3	5.6
Description	6.1	5.2	5.4	5.6
Overall	7.5	4.9	4.3	5.6

Table 2. *The lengths of the recall protocols as a proportion of the length of the presented passage for each sort of passage and version in Experiment 1*
Each entry is summed over 24 subjects and 3 passages.

	Original	Revised	Random	Overall
Story	0.71	0.43	0.30	0.48
Description	0.32	0.30	0.32	0.32
Overall	0.52	0.37	0.31	0.40

prehensibility, there was a reliable effect of version on the memorability of the stories (minF' = 18.31, d.f. = 2,8, $p < 0.01$): the original stories were better remembered than the revised random versions (minF' = 20.66, d.f. = 1,8, $p < 0.01$), and both these versions were better remembered than the simple random versions of the stories (minF' = 15.95, d.f. = 1,8, $p < 0.01$). The trend over the three versions was again significant (Page's $L = 329$, $z = 5.92$, $p \ll 0.001$). But there were no reliable effects of version on the memorability of the descriptions.

As we predicted, the original versions of the stories were easier to understand and to recall than randomised versions, but the restoration of referential continuity reliably ameliorated the effects of randomisation. In the case of the descriptive passages, which had a more diffuse continuity, the effects of randomisation were negligible whether or not referential continuity had been restored.

The possibility of profiting from referential continuity requires an ability to make the necessary 'bridging' inferences (see, e.g. Clark, 1977). Hence, subjects who are poor at making these inferences should show less benefit from the restoration of continuity in a randomised story than subjects who are good at making them. In order to test this prediction, we carried out a study of children's memory for stories presented in different versions.

Experiment 2

Little is known about children's developing awareness of the structure of stories, though work by Poulson *et al.* (1979) has demonstrated the beginnings of such awareness in children as young as 4 years. Previous unpublished studies by the second author have established that skilled and less skilled comprehenders did not differ in *verbatim* retention ot digits and short sentences, but they do differ in their propensity to make inferences. In one study, for example, two such groups were given a recognition test for sentences that they had heard. The skilled comprehenders were more likely to make recognition errors based on plausible inferences from the passages than were the less skilled comprehenders; the skilled comprehenders, as one would expect, were better than the less skilled comprehenders in recognising the original sentences and in rejecting confusion items based on implausible inferences (see Oakhill, in press). Kintsch and Vipond (1979) have similarly suggested that inferential ability may be an important factor in coping with incoherent discourse, and Warren *et al.* (1979) emphasise the role of inferential connections in the derivation of story structure.

In the present experiment, two groups of 7--8-year-olds were selected from a tested sample of 168 children. The two groups were matched on tests of word recognition and reading vocabulary, but they differed in their performance on a reading comprehension test. Both were asked to recall normal, random and revised random stories. We predicted that the skilled comprehenders would benefit more than the less skilled comprehenders from the restoration of referential continuity in the randomised stories, since the skilled comprehenders would be able to exercise their greater inferential skills in carrying out the required 'bridging' inferences. We also predicted that the skilled comprehenders would remember the original stories and the revised random stories better than would the less skilled comprehenders, but that there would be no difference between the groups' memory for the random stories.

Method

Subjects

Twenty four children from two Brighton primary schools participated in the experiment. They were divided into two groups that differed in comprehension ability. Twelve less skilled comprehenders were chosen according to the following criteria: their reading accuracy age was above or equal to their chronological age, but their comprehension age was below their chrono-

logical age and was at least half a year below their reading accuracy age. Twelve skilled comprehenders were chosen who were matched with the less skilled group for gender (7 boys and 5 girls in each group), chronological age (the mean for both groups was 7.9 years), and word recognition age on the *Neale Analysis of Reading Ability* (the mean for the skilled group was 8.4 years, and for the less skilled group 8.5 years). The two groups were also selected so that their scores on a sight—recognition vocabulary test (an adapted form of the *Gates–McGinitie Primary Two Vocabulary Test*) did not differ significantly. They did differ in comprehension ability, as measured by the *Neale* test, however, with no overlap between the groups (the mean for the skilled group was 9.1 years, and for the less skilled group was 7.3 years).

Materials

Six passages, each five sentences in length (48–50 words), were written in a suitable vocabulary and based on story topics which were familiar to children. In each passage the main person and object occurred in the first sentence and were pronominalised thereafter, except in the fourth sentence where the object was reintroduced and in the fifth sentence where the person was reintroduced. We constructed a second version of each passage in which the sentences occurred in a random order but were otherwise unchanged. A different random order was used for each passage. We then constructed a third version by replacing the noun phrases of the randomised passage, where necessary, so as to re-establish the referential continuity of the story. Each passage was typed on a separate card using a typeface that was easily legible for children. A typical example of a story in its three versions is presented below:

> David was playing with his big, coloured ball in the garden. He bounced it so hard that it went right over the fence. The people next door were out so he climbed over to get it. He found his ball and threw it back. David carried on with his game.
>
> He found his ball and threw it back. The people next door were out so he climbed over to get it. David carried on with his game. He bounced it so hard that it went right over the fence. David was playing with his big coloured ball in the garden.
>
> David found his big coloured ball and threw it back. The people next door were out so he climbed over to get it. He carried on with his game. He bounced his ball so hard that it went right over the fence. David was playing with it in the garden.

Design

The subjects acted as their own controls and read and recalled two stories in each of the three experimental versions: original, random and revised.

The passages were put into pairs that each contained one passage about a female subject and one about a male. For a given subject, the pair of passages was presented in the same version, one passage after the other. The pairs appeared once in each version in every block of three subjects, producing three sets of materials. Three different orders of presentation of the three versions were also used to control for practice effects between versions. In this way each passage version appeared equally often in each presentation position. The occurrence of sets of materials within presentation orders was counterbalanced, as far as possible, over the twelve subjects in each group.

Procedure

The children were familiar with the experimenter from the previous test sessions. They were seen individually in a quiet room. At the beginning of the experiment, they were told that they would be given six short stories to read, and, after they had read each story, they would have to repeat as much of it as they could remember. They were told to try to recall the exact words of the story but, if they were unable to do so, they should put what they could remember into their own words.

After the children had read each story, the instructions were repeated in full in order to eliminate any recency effects. The children had as much time as they wanted to recall the stories: the next story was not presented until they had said that they had finished their recall.

Results

Six independent adult judges divided each story into nine 'idea units', and the children's recall protocols were scored in relation to those units—for the most part simple surface clauses—on which at least four of the judges were agreed. Table 3 presents the mean numbers of idea units recalled by the skilled and less skilled comprehenders for the three sorts of passage. A child was deemed to have remembered an idea unit provided that its gist was correct, i.e. synonymous terms were treated as correct, and there was no penalty for omitting adjectival or adverbial modifiers, articles and other non-essential items. The skilled comprehenders recalled more ideas overall than did the less skilled comprehenders ($F = 13.93$, d.f. = 1,22, $p < 0.005$); the version in which a story occured had a significant effect on the number of ideas recalled ($F = 16{,}51$, d.f. = 2,44, $p < 0.001$); and there was an interaction between reading ability and the version of the story ($F = 3.25$, d.f. =

Table 3. *The mean number of 'idea units' (max. = 9) recalled by the skilled and less skilled readers for the three sorts of story in Experiment 2*

	Original	Revised	Random	Overall
Skilled	7.3	6.1	4.8	6.1
Less skilled	5.2	3.5	3.9	4.2
Overall	6.2	4.8	4.4	5.2

2,44, $p < 0.05$). Since we were interested in all the comparisons in performance, we made use of the Newman–Keuls procedure. The tests revealed that skilled comprehenders recalled more of the original stories than the revised random stories ($p < 0.05$), and more of the revised random stories than the unrevised random stories ($p < 0.05$). The less skilled comprehenders recalled more of the original stories than the revised random stories ($p < 0.05$), but they did *not* recall reliably more of the revised random stories than of the unrevised random ones. The skilled comprehenders recalled more of the original and revised random stories than did the less skilled comprehenders ($p < 0.05$), but there was no reliable difference between the two groups in their ability to recall the unrevised random stories. An analysis of the mean numbers of words recalled yielded substantially the same pattern of results.

In summary, both predictions were confirmed. Indeed, only the skilled comprehenders were able to take advantage of the restoration of referential continuity in randomised stories. Their superior abilities helped them with remembering the original and the revised random stories. But, if the task was to recall a jumbled story in which there was no referential continuity, it made little difference whether a subject was a skilled or less skilled comprehender.

General discussion

Our basic findings can be illustrated by considering the following brief story, derived from Rumelhart (1975), and which we used in Experiment 2:

> Jenny was holding on tightly to the string of her beautiful new balloon. She had just won it and was hurrying home to show her sister. Suddenly, the wind caught it and carried it into a tree. The balloon hit a branch and burst. Jenny cried and cried.

When the order of the sentences is randomised, the passage is rendered incoherent:

> She had just won it and was hurrying home to show her sister. Suddenly, the wind caught it and carried it into a tree. Jenny was holding on tightly to the string of her beautiful new balloon. Jenny cried and cried. The balloon hit a branch and burst.

It ceases to be clear quite what the story is about—what it is referring to. The first sentence refers to 'she' and to 'it', and the reader is likely to imagine a girl and some sort of prize. Later, when it is asserted that Jenny was holding on to the string of her new balloon, it seems that reference is being made to a new person with a new possession. Such misreadings are ruled out by restoring referential continuity in the randomised sentences:

> Jenny had just won a beautiful new balloon and was hurrying home to show her sister. Suddenly the wind caught it and carried it into a tree. Jenny was holding on tightly to the string of her balloon. She cried and cried. It hit a branch and burst.

The text is coherent, but now the sequence of events is slightly odd. Jenny seems to have been holding on to the string of the balloon *after* it was carried into the tree. Some readers might construe the sentences as containing the wrong auxiliary: it would make better sense if it had read, 'Jenny *had been* holding on tightly to the string...'. Nevertheless, the general impression created by such texts is of a somewhat implausible sequence of events. The results of our experiments confirmed that both coherence and plausibility exert an effect on the comprehensibility and the memorability of stories. Purely descriptive passages whose coherence is less dependent on referential connections between one sentence and the next can withstand randomisation with very much less effect on performance. Readers who are poor at making the inferences required to establish coreference between sentences gain relatively little from the re-establishment of coherence in randomised stories.

Why do the different versions of the stories differ in their memorability? One answer might be that the differences are an immediate consequence of comprehensibility. But, the original stories with their plausible sequences of events are also more redundant than the implausible sequences of the revised random versions, and general knowledge can aid the readers to make good any deficiences in their memories of redundant sequences. What still remains to be explained, however, is the greater memorability of the revised random stories over the unrevised versions. A number of factors may be at work. Incoherent texts are likely to be interpreted, if at all, as refering to a greater number of different entities than do their coherent counterparts—a factor that we illustrated in the analysis of the story about the girl and the balloon. Likewise, a coherent text yields a unitary representation that integrates the information from a number of sentences, whereas an incoherent

text yields only independent fragments of an interpretation. Integrated representations are more resistant to forgetting, and allow a greater degree of 'elaboration' (Craik and Tulving, 1975). We can be sure, however, that story grammars have little role to play in accounting for the better recall of revised random texts. The sequence of events in such texts is outside the restricted and stereotyped domain of a story grammar.

The theory of interpretation that comes closest to answering our questions has been put forward by Kintsch (1974) and Kintsch and van Dijk (1978). Kintsch argues that the structure of a text is defined in terms of the relations between the basic propositions that it expresses. There is, of course, no algorithm that will deliver such propositional representations, and Kintsch and van Dijk point out that they have no account of how a list of propositions is derived from a text. They are concerned with the processes by which a referentially coherent microstructure is built up from the basic propositions, and with the processes that then lead to the construction of a single tree that represents the macrostructure of a text in much the same way as a tree generated by a story grammar. They argue that referential coherence is established by operators that look for antecedents for the expressions in the proposition under analysis, and, if no plausible antecedents can be found, then various inferential mechanisms are brought into play. Their essential claim is that referential integration can occur provided that there is an overlap between the arguments of the proposition under analysis and the arguments of those propositions that have already been integrated.

The main difficulty with Kintsch and van Dijk's theory arises from the propositional representations on which it is based. An assertion such as:

The fat man drinks beer.

is supposed to be mentally represented by a corresponding string of symbols:

(DRINKS MAN BEER) and (FAT MAN)

There are many technical difficulties with the details of the notation, e.g. the representation of the scope of quantifiers. But, granted that they could be fairly easily solved (see e.g., the papers in Part 1 of Findler (1979) for some solutions in the closely related framework of associative network theory), it still remains clear that the mere repetition of an argument, such as MAN or BEER, does not guarantee that the same entity is being talked about, e.g., 'the fat man drank a beer and so did another fat man standing next to him'. It is doubtful whether symbols corresponding to words provide a rich enough domain for establishing referential relations. In common with a number of theorists, we believe instead that discourse is ultimately inter-

preted in the form of a model which is specific to the particular discourse and in which there are tokens that correspond directly to referents (see Karttunen, 1976; Stenning, 1978; Webber, 1978; Johnson-Laird, 1978). We further believe that explicit rules can be given for the construction of such models from text. As Johnson-Laird and Garnham (1980) have argued, the tokens in a discourse model provide the set of possible referents for anaphors in the text. There is indeed evidence that when subjects are unable to form a mental model, because a description of a spatial layout is grossly indeterminate, their memory for the gist of the description is much poorer though they may well be better able to remember a number of *verbatim* details (Mani and Johnson-Laird, in press). Such results naturally suggest that a superficial representation containing information about the surface form of a text may be a precursor to a richer representation in the form of a mental model. It is conceivable that the superficial representation is something like the list of propositions described by Kintsch and van Dijk, though another suggestion is that it is basically phonological. The same two levels of representation, superficial and mental model, may have occurred with the texts in our experiments: a randomised story may be represented only as a sequence of more or less independent 'propositions', whereas once referential continuity has been restored a model of the discourse may be constructed.

In conclusion, although there is no good evidence that comprehension depends on some specific knowledge of the structure of stories, which is independent of their content, it clearly does depend on a vast amount of knowledge. There are great difficulties in formalising this knowledge, as workers in psychology and artificial intelligence have found out to their cost. On the one hand, story grammars are an attempt to factor out one part of this knowledge, and the basic idea behind them is a good one. However, there are so many different kinds of texts that story grammars inevitably seem simplistic, and they may prove to be redundant in the light of accounts of the other sorts of knowledge underlying the comprehension of discourse. On the other hand, the problem of coherence of text—its microstructure—is more tractable. Our results imply that referential coherence is a pre-eminent factor in the interpretation of discourse: only when it is present can a reader or listener begin to construct a unitary model of the discourse.

References

d'Aulaire, E. and d'Aulaire, P. O. (1979) The wonderful anti-ant machine. *International Wildlife*, *9* (Dec.), 24–29.

Black, J. B. and Wilensky, R. (1979) An evaluation of story grammars. *Cog. Sci.*, *3*, 213–230.

Bransford, J. D. and McCarrell, N. S. (1975) A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In W. B. Weimar and D. S. Palermo (eds.), *Cognition and the Symbolic Processes*. Hillsdale, NJ, Erlbaum.

Chomsky, N. (1957) *Syntactic Structures*. The Hague, Mouton.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. Cambridge, Mass., MIT Press.

Clark, H. H. (1977) Bridging. In P. N. Johnson-Laird and P. C. Wason (eds.), *Thinking: Readings in Cognitive Science*. Cambridge, Cambridge University Press.

Craik, F. I. M. and Tulving, E. (1975) Depth of processing and the retention of words in episodic memory. *J. of exper. Psychol. Gen.*, *104*, 268–294.

Cure, K. (1980) Welcoming the year of the monkey. *Travel/Holiday*, *153* (1) 60–63.

Ehrlich, K. and Johnson-Laird, P. N. (1980) Spatial descriptions and referential continuity. Mimeo, Laboratory of Experimental Psychology, University of Sussex.

Findler, N. V. (1979) (ed.) *Associative Networks: Representation and Use of Knowledge by Computers*. New York, Academic Press.

Garrod, S. and Sanford, A. J. (1977) Interpreting anaphoric relations: The integration of semantic information while reading. *J. verb. Learn. verb. Behav.*, *16*, 77–90.

Gazdar, G. J. M. (In press) Phrase structure grammar. In G. K. Pullum and P. Jacobson (eds.), *The Nature of Syntactic Representation*. London, Croom Helm.

Greenway, J. (1965) (ed.) *The Primitive Reader*. Hatsboro, PA, Folklore Associates.

Grice, H. P. (1975) Logic and conversation. In P. Cole and J. M. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*. New York, Seminar Press.

Haviland, S. E. and Clark, H. H. (1974) What's new? Acquiring new information as a process in comprehension. *J. verb. Learn. verb. Behav.*, *13*, 512–521.

Johnson-Laird, P. N. (1978) Mental models of meaning. Paper delivered to Sloan Workshop on Computational Aspects of Linguistic Structure, University of Pennsylvania. To appear in A. K. Joshi, I. A. Sag and B. L. Webber (eds.), *Elements of Discourse Understanding*. Cambridge, Cambridge University Press.

Johnson-Laird, P. N. and Garnham, A. (1980) Descriptions and discourse models. *Ling. Phil.*, *3*, 371–393.

Karttunen, L. (1976) Discourse referents. In J. D. McCawley (ed.), *Syntax and Semantics 7: Notes from the Linguistic Underground*. New York, Academic Press.

Kintsch, W. (1974) *The Representation of Meaning in Memory*. Hillsdale, NJ, Erlbaum.

Kintsch, W. and van Dijk, T. A. (1978) Towards a model of text comprehension and reproduction. *Psychol. Rev.*, *85*, 363–394.

Kintsch, W. and Vipond, D. (1979) Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (ed.), *Perspectives on Memory Research*. Hillsdale, NJ, Erlbaum.

Levelt, W. J. M. (1974) *Formal Grammars in Linguistics and Psycholinguistics Vol. II: Applications in Linguistic Theory*. The Hague, Mouton.

Mandler, J. M. and Johnson, N. J. (1977) Remembrance of things parsed: Story structure and recall. *Cog. Psychol.*, *9*, 111–151.

Mandler, J. M. and Johnson, N. J. (1980) On throwing out the baby with the bathwater: A reply to Black and Wilensky's evaluation of story grammars. *Cog. Sci.*, *4*, 305–312.

Mani, K. and Johnson-Laird, P. N. (In press) The mental representation of spatial descriptions. *Mem. Cog.*

Miller, G. A. (1967) Project grammarama. In *The Psychology of Communication: Seven Essays*. London, Allen Lane.

Oakhill, J. V. (In press) Constructive processes in skilled and less skilled comprehenders' memory for sentences. *Br. J. Psychol.*

O'Sullivan, S. (1966) (ed.) *Folktales of Ireland.* Chicago, Chicago University Press.

Poulson, D., Kintsch, E., Kintsch, W. and Premack, D. (1979) Children's comprehension and memory for stories. *J. exper. Child Psychol., 28,* 379–403.

Putigny, B. (1976) *Easter Island.* New York, The Two Continents Publishing Group.

Rumelhart, D. E. (1975) Notes on a schema for stories. In D. G. Bobrow and A. M. Collins (eds.), *Representation and Understanding.* New York, Academic Press.

Stenning, K. (1978) Anaphora as an approach to pragmatics. In M. Halle, J. Bresnan and G. A. Miller (eds.), *Linguistic Theory and Psychological Reality.* Cambridge, MA, MIT Press.

Thorndyke, P. W. (1977) Cognitive structures in comprehension and memory of narrative discourse. *Cog. Psychol., 9,* 77–110.

Warren, W. H., Nicholas, D. W. and Trabasso, T. (1979) Event chains in understanding narratives. In R. O. Freedle (ed.), *New Directions in Discourse Processing.* Norwood, NJ, Ablex.

Webber, B. L. (1978) *A Formal Approach to Discourse Anaphora.* BBN Tech. Rep. No. 3761. Cambridge, MA, Bolt, Berenek and Newman.

Résumé

Deux expériences ont été faites pour rechercher le rôle de la continuité référentielle dans la compréhension du discours. Dans la première expérience, un groupe d'étudiants écoute des histoires et des descriptions présentées sous trois versions différentes: la version originale, la version où l'ordre des phrases est aléatoire et la version dans laquelle la continuité reéférentielle a éte restaurée par le remplacement des prénoms et autres termes par des syntagmes nominaux appropriés et plus complets. Les sujets mémorisent mieux les histoires originales et les trouvent plus compréhensibles que les versions aléatoires. Cependant, la restauration de la continuité référentielle améliore les effets dûs à la présentation aléatoire. Les descriptions, qui ont peu de continuité référentielle d'une phrase à l'autre sont, comme prévu, moins sensibles à la présentation en ordre aléatoire. Dans l'expérience 2, un groupe de sujets compétents et un groupe de sujets peu compétents sont sélectionnés parmi des enfants de 7–8 ans. On suppose que la différence entre les groupes provient essentiellement de leur capacité d'inférence dans la lecture du texte. Les deux groupes lisent une série d'histoires courtes présentées selon les trois modalités précédentes. Ainsi que prévu, les effets améliorant la restauration mnémonique de la continuité référentielle dans une histoire présentée en ordre aléatoire ne se retrouvent que dans le groupe le plus compétent. Les auteurs discutent de ces résultats en relation avec les théories de la grammaire des histoires, de la microstructure du texte et des modèles mentaux du discours.