Artificial Intelligence Review **15:** 295–323, 2001.

Representation, Coherence and Inference

ELLIOT SMITH and PETER HANCOX
School of Computer Science, University of Birmingham, Edgbaston, B15 2TT, UK

Abstract. Approaches to story comprehension within several fields (computational linguistics, cognitive psychology, and artificial intelligence) are compared. Central to this comparison is an overview of much recent research in cognitive psychology, which is often not incorporated into simulations of comprehension (particularly in artificial intelligence). The theoretical core of this experimental work is the establishment of coherence via inference-making.

The definitions of coherence and inference-making in this paper incorporate some of this work in cognitive psychology. Three major research methodologies are examined in the light of these definitions: scripts, spreading activation, and abduction.

This analysis highlights several deficiencies in current models of comprehension. One deficiency of concern is the 'one-track' behaviour of current systems, which pursue a *mono-stratal* representation of each story. In contrast, this paper emphasises a view of adaptive comprehension which produces a 'variable-depth' representation. A representation is pursued to the extent specified by the comprehender's goals; these goals determine the amount of coherence sought by the system, and hence the 'depth' of its representation. Coherence is generated incrementally via inferences which explain the co-occurrence of story elements.

Keywords: Cognitive modelling, coherence-driven processing, comprehension metrics, variable-depth representations, story comprehension

1. Introduction

Many models of story comprehension exist in the computational linguistics and artificial intelligence (AI) literature. Often, systems based on these models produce convincing results, such as accurate summaries. However, the mechanisms which lead to these results are often inadequate as *cognitive* models, even in those cases where cognitive plausibility is claimed. This inadequacy is often a consequence of ignoring experimental evidence from cognitive psychology, which focuses on a fine-grained appraisal of the nature and purpose of comprehension. This evidence both confronts the implications of specific models (such as those based on scripts) and emphasises the *flexibility* of comprehension, a factor often missing from implemented systems.

In this paper, we collate and compare some of this recent research on story comprehension in computational linguistics, cognitive psychology, and artificial intelligence. Firstly, we examine the following three features of comprehension:

1. The representations produced during comprehension.
2. How coherence is implicated in the production of representations.
3. The inferences which generate coherence in the representation.

We then describe some previous story comprehension models devised by researchers in AI, with the aim of evaluating these models in the light of the above features, simultaneously demonstrating how they diverge from psychological evidence. Our main emphasis is on high-level, executive processes (for example, making decisions about coherence requirements and evaluation of competing interpretations).

Finally, we make several suggestions about how to compensate for inadequacies in current AI models.

2. What Does it Mean to Understand a Story?

Recent research has shown that when people read stories, comprehension is primarily *explanation-driven* (Graesser et al., 1994). In practice, this means that comprehenders look for answers to 'Why?' questions, as opposed to 'What happens next?', 'How?', 'Where?', or 'When?' questions. The product of the comprehension process is a set of explanations of why particular information has been included in the story. Explanations are 'backward oriented in narrative time, and serve to unit the focal sentence with either text information or prior-knowledge-based inferences' (Trabasso and Magliano, 1996); that is, explanation moves from the eventualities explicitly stated in the story to hypotheses that explain how the eventualities relate to each other.

It may seem obvious that comprehension is explanation-driven, but other researchers have suggested that comprehension is *expectation-driven*. Expectation-driven comprehension is based on predictive inferences which 'fill' uninstantiated parts of a memory structure (Schank and Abelson, 1978). In expectation-driven comprehension, the comprehender typically makes many predictions about the forthcoming story, which are either supported or denied by that text. However, the psychological literature indicates that predictive inferencing is limited in scope, and highly constrained by the content of the text read so far (Trabasso and Magliano, 1996; see section 4.3).

The set of explanations derived from the statements of the story, along with semantic translations of the explicit statements, are stored as a representation in episodic long-term memory. Story representations are then accessible for the purpose of connecting new story statements, producing summaries, creating new items in semantic memory (learning), and so on. The ability to produce a representation is often used as the litmus test of a comprehension system; the representation is usually accessed via summarisation and/or question answering (e.g. DeJong, 1979; Alterman, 1985; Alterman and Bookman, 1990; Lehnert et al., 1983). (However, these access operations can often add extra layers of uncertainty to system evaluation: the production of answers and summaries can itself alter the representation.)

Note that a representation will never contain all of the possible information that could be generated from a story: the comprehension process creates a representation that is a subset of the *possible explanations space* of the story being read. A comprehender has two principal mechanisms for pruning the possible explanations space, so that only plausible and relevant information is retained in their representation (Norvig, 1989):

– *Application of knowledge sources*
  The comprehender's knowledge sources delineate the plausible explanations for a given story. These knowledge sources represent the comprehender's *competence* in the field of story comprehension (as opposed to their actual performance) (Crystal, 1971). Knowledge sources contain such information as the way events occur in causal sequences in the real world (Mandler, 1984), and the way stories are typically structured (Correira, 1980).
– *Consultation of comprehension goads*
  The comprehender typically has a set of goals that determines which of the plausible explanations are relevant to the task at hand. For example, if a detailed representation is required (e.g. the comprehender knows they will be tested on the content of the text), then the goal may be to include many possible explanations so that this goal is fulfilled.

A typical pattern of 'pruning' the possible explanations space is shown diagrammatically in Figure 1.

A representation can be viewed as a summary (or subset) of the possible explanations space. According to Alterman (1991), 'The capacity to summarise is a fundamental property of intelligence'. It allows a comprehender with limited processing and storage capacity to extract relevant and interesting information from a potentially unbounded explanations space. This notion is implicit in many story comprehension systems, but rarely explicitly stated.

2.1. *Characteristics of human story representations*

Pruning of the possible explanations space according to available *knowledge sources* and current *goals* produces a representation which is not a mere copy of the information in the story. Instead, the representation is edited in line with

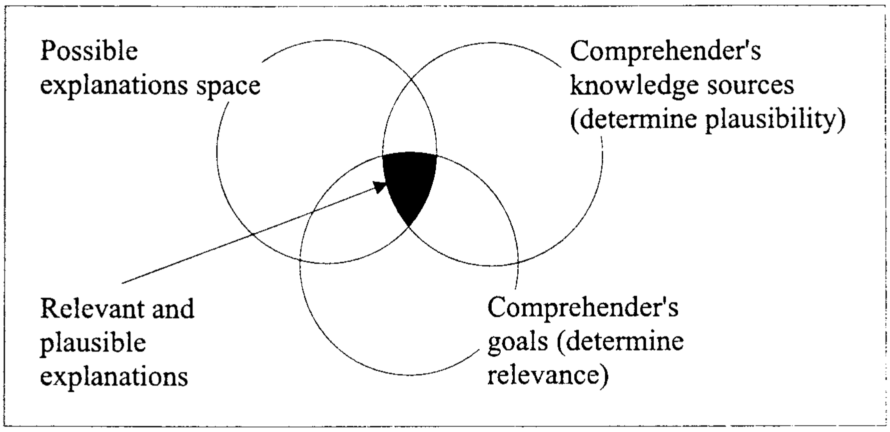

Figure 1. Pruning of the possible explanations space

these two sources of control, resulting in the manifestation of the following qualities:

– *Addition, deletion, and reorganisation*

  People add events and states that weren't explicitly described in the story to their representations of stories (Mandler and Johnson, 1977). These events may be at the micro-level (the addition of individual story statements), or at a generalised macro-level (the addition of new macro-statements) (see van Dijk, 1977 and section 2.3.3). People also delete information that they consider irrelevant.

  The information included in a representation is often reorganised in line with the comprehender's knowledge about how event sequences occur in the world. For example, if a story seems to belong to a generic category (such as 'the fairy-tale'), but its event sequence veers from the norm for this genre, then the representation tends to be reordered in line with the expected event sequence (Mandler and Johnson, 1977; Thorndyke, 1977).

– *Explicated causal relations*

  People include causal relationships between eventualities in their representations. These relations make the representation coherent, tying eventualities at one end of the story to eventualities at the other. For example, intentional actions are explained by reference to the superordinate goals of characters (Graesser et al., 1994).

– *Variable-depth*

  People can produce representations at different levels of specificity. Various parameters, such as working memory capacity (Whitney et al., 1991), domain knowledge (Noordman and Vonk, 1992), and reader goals (Graesser et al., 1994) influence the extent of comprehender's representation.

The competence underlying these changes to the representation resides in the knowledge sources; for example, new events may be added if a schema has been partially matched by an input story. The performance aspect of these changes resides in the comprehender's goals; for example, the decision to add new events by instantiating the partially-matched knowledge source depends on the comprehender's drive to integrate new information into their representation. However, even with the two sources of pruning, there may still be cases where two or more representations are in competition. In this case, comparison of their respective qualities is required, as described in the next section.

2.2. *How to measure the quality of a representation?*

It is clear that there are differences between what people read, and what people remember of what they read. How can a comprehender distinguish between differences which are desirable, and those which are superfluous? Alterman defines several desirable properties of a summary (Alterman, 1991); these properties are also generally applicable to any form of representation. These desirable properties are that a representation should:

– *Include important events*
  A representation is only useful to the extent that it includes the 'important' events in a story. The most important events can be defined as those that are retained in the most abstract version of the representation (see section 2.3.3).
– *Be coherent*
  A representation should 'hold together and make sense' (Alterman and Bookman, 1990). While this is easy to say, it is much harder to specify. One of the main ways in which coherence is established is by linking event sequences together via various kinds of relation, of which causal relations seem to be the most important. Another way is extraction of the 'gist(s)' of a story, by finding one or more macro-events that summarise it (see section 3.2 for more information about this distinction).
– *'Cover' the content*
  The representation should include, either explicitly or implicitly, all of the events in the story. Any of the explicit events of the story that are left out of the representation should be recoverable from it, via further inference processes.

These requirement characterise what an *ideal* representation should look like. They also describe the kinds of information that a comprehender should be looking for while reading a story. That is, they should attempt to produce coherent explanations, that both implicitly cover the events of the story and highlight the most important events. By examining the relative acceptability in terms of these requirements, decisions between competing representations can be made.

2.3. *What kinds of representations are produced?*

The representation produced by a comprehension model is usually a set of nodes, and a set of links between them. Each node represents a single explicit or implicit story eventuality. The links may be local between pairs of nodes, or may be high-level, global relations between clusters of nodes. The distinction between low-level, local links, and high-level, global links determines the kind of coherence captured by the link (see section 3.2 for more details). In the psychological literature, a representation consisting of a set of nodes and a set of links is often called a *propositional textbase* (Kintsch and van Dijk; Garnham, 1987).

Within this broad framework, there have been numerous theories about the exact structure of the representation produced. In simple terms, these theories attribute different amounts of local and global connectivity to their representations. For example, the causal network approach emphasises global relations which cluster a whole group of nodes.

Simplifying slightly, there are three principal forms of representation that have been suggested for stories: *linear, heterarchical*, and *hierarchical*. These are covered in the next three sections.

2.3.1. *Linear representations – scripts and the causal chain*

The most widely-used of the linear representations is the causal chain, produced by script-based processing (Schank and Abelson, 1978); several models employing script-based processing have been implemented as programs (Lehnert et al., 1983; Schank and Wilensky, 1978).

Script-based processing is based on inferring categories for event sequences. A causal chain is a sequence of events which remains constant for any instance of a particular category and consists of a set of nodes with pairwise causal connections between them. For example, in the $SUBWAY script, a $PATRON's ticket-buying enables them to pass through the barrier, which in turn enables them to board the train, which causes their movement from one place to another (Dyer, 1992). The representation produced during comprehension is an instantiated causal chain. Most researchers assume that a script-based processor can only identify a single antecedent and a single consequent for each event (but see section 2.3.3 for a critique of this assumption).

Schank and Abelson distinguish five types of link used to instantiate a causal chain:

1. Actions can result in state changes.
2. States can enable actions.
3. States can disable actions.
4. States or actions can initiate mental states.
5. Mental states can be reasons for actions (or, in the other words, mental states can motivate actions).

It is interesting to note the similarity between these types of causality and those employed by story grammarians (e.g. Rumelhart, 1975; Mandler and Johnson, 1977), and causal network theorists (see section 2.3.2).

2.3.2. *Heterarchical representations – causal networks*

Heterarchical network representations were devised in response to what were seen as inadequacies in the causal chain approach (Trabasso and van den Broek, 1985). They are based on the idea that a single eventuality may have multiple antecedents and consequents, and that causal strengths are not absolute but depend on sufficiency and necessity conditions, as well as temporal order. To capture these features of the representation, links may connect non-adjacent eventualities and multiple links may enter or leave a node.

Comprehension results in a network representation, rather than a single causal chain that ignores peripheral events. The nodes of the network represent individual eventualities in the story, such as character states or actions. The links between the nodes represent causal dependencies between nodes. A pair of nodes and a link of the form $A \rightarrow B$ are interpreted as *A causes B*.

A causal link between two nodes is not guaranteed and does not have an absolute strength. Given two eventualities. *A* and *B*, there are several constraints that influences whether a causal link between them can be inferred (Trabasso and van den Broek, 1985; Trabasso et al., 1989; Trabasso and Magliano, 1996; van den Broek and Trabasso, 1986; van den Broek, 1990a; van den Broek, 1990b; van den Broek, 1994):

– *Temporal priority*
  This constraint can be summarised as ‘the causing event must always happen before its effect event’, or ‘a cause never occurs after its consequent’ (van den Broek, 1990b).
– *Operativity*
  This means that a cause must be active at the time when its effect occurs. For example, for ‘John met Wendy’ to cause ‘John married Wendy’, the state of ‘John knowing Wendy’, brought about by John meeting Wendy, must be operative at the instant before their marriage occurs.

- *Necessity*
  This constraint is one of logical necessity, and is tested for counterfactually be applying the following rule to *A* and *B*: ‘It *A* had not occurred in the circumstances of the story, the *B* would not have occurred’ (Trabasso et al., 1989). For example, ‘John picked up a knife’ is a necessary condition for ‘John stabbed Peter with the knife’: if John hadn’t picked the knife up, then the stabbing with the knife wouldn’t have occurred.
- *Sufficiency*
  This constraint is weaker than that of necessity, and is tested for by applying the following rule to *A* and *B*: ‘If *A* occurs in the circumstances of the story, then *B* is likely to occur’ (van den Broek, 1994). For example, ‘John struck a match’ is a sufficient condition for ‘John lit his cigarette’, providing the two events are temporally proximal, and John is holding a cigarette when the striking occurs.

2.3.2.1. *Types of causal link*. The constraints in the previous section can be used to determine the type of causal link between two eventualities. Trabasso et al. identify four types of causal link: physically causes (ϕ), psychologically causes (ψ), motivates (*M*), and enables (*E*) (Trabasso et al., 1989).

The type assigned to a link depends on how the constraints outlined above interact with the contents of the eventualities. An example rule for creating a causal link is as follows:

> *A motivates B [M (A, B)]* if *A* is temporally prior to *B* and *A* is operative when *B* occurs and *A* is necessary for *B* and *A* contains goal information.

Example:
(a) Billy wanted a bike. (b) He went to the shop to buy one. → *M (a, b)*

This formulation of criteria for the existence of a causal link, plus the description of different kinds of links, makes this theory finer-grained than theories which equate causal paths with pre-determined, set-in-stone sequences of patterns (c.f. scripts). However, implementations of causal networks are currently limited or pseudocode only (e.g. van den Broek, 1990a).

2.3.2.2. *A note on causal chains and causal networks*. The causal network is often cited as an alternative to the causal chain, in that it allows the representation of multiple antecedents and consequents for an eventuality; causal chains supposedly do not (van den Broek, 1994). However, a glance at Schank and Abelson’s book about scripts provides an example causal chain that includes a single state with multiple consequents: a chair’s state change causes someone to fall off it, a beer to spill, and the creation of a noise (Schank and Abelson, 1978). Also, although a causal chain is generally seen as being linear, scripts also provide a limited hierarchy: a script can be seen as a higher-level chunk (or macrostructure) containing or 'covering' the events in the causal chain.

2.3.3. *Hierarchical representations – story trees*

A third common form of representation posited by researchers is hierarchical. The earliest models to emphasise hierarchical structure were *story grammars* (Rumelhart, 1975). In this representation scheme, there are still pairwise 'linear' links between nodes, as in heterarchical representations. In this sense, the representation is *locally* linear. However, the network is also extended 'upwards': nodes may be introduced at a higher-level of the network, which have links to nodes on the lower levels. It may be easier to imagine this as a collection of networks, arranged in a pyramid of tiers. The higher-level tiers are smaller as they contain fewer nodes; these nodes generalise information in the lower layers (van Dijk, 1977).

A second definition of hierarchy is the *macrostructure*, which describes 'global levels of descriptions' (van Dijk, 1977). A macrostructure 'defines the most important or essential object or event denoted by a sequence of propositions' (van Dijk, 1977). Correira (1980) implemented a model in which macrostrucutres are equated with nodes in a story grammar tree: 'a macrostructure is a node in a story tree whose immediate descendants consist of the subordinate propositions by which the node is implied'. A macrostructure is thus entailed by the propositions it covers, and its semantic representation is 'a function of the meaning and reference of the constituent propositions of the explicit text base and the relations between those propositions' (van Dijk, 1977).

However, each node in a hierarchical story representation tends to have a single antecedent and/or consequent, leaving these theories open to the criticisms levelled at causal chain theories. As causal network theorists explain, this may not be the most accurate way of describing relations between eventualities. For example, a single eventuality encapsulated in a macrostructure may cause multiple consequent eventualities.

2.3.3.1. *A Note on Story Grammars*. In spite of their popularity as theories of story comprehension in the 1970s, story grammars fell out of favour with AI researchers in the early 1980s. Several researchers highlighted their inability to cope with embedded stories (Black and Wilensky, 1979); some attacked story grammarians' failure to distinguish between a grammar and a parser, resulting in muddled accounts of processing (Garnham, 1983). However, models suggested as possible replacements were often themselves subject to the same criticisms: as Frisch and Perlis (1981) point out, any formal system ('including all knowledge-based, plan-driven inference systems') can be expressed as a set of rewrite rules. Complaints about mathematical properties of story grammars thus have little bearing on their outputs (story *trees*) as feasible models of story *representation*; rather, such complaints highlight the more general problem of constructing formal grammars for certain classes of structure.

More importantly, in the context of *story trees as representations*, were doubts about how well they accounted for evidence from story recall experiments. The 'depth' of a particular element in a story tree was taken as a measure of its probability of being recalled: the higher the element in the tree, the more likely it would be recalled (Mandler and Johnson, 1977). Later research seemed to show that it was causal connectivity rather than hierarchical position which influences probability of recall, although the results were somewhat inconclusive (van den Broek and Trabasso, 1986). As a result, story tree representations remain important, particularly in their wider capacity as hierarchies of varying abstraction.

3. Coherence

The distinctions made in section 2 between linear, heterarchical and hierarchical representation schemes depend on the establishment of different kinds of coherence. A representation extends into a particular dimension if relations can be established within that dimension.

A distinction commonly made in the psychological literature is between *local* and *global* coherence (Graesser et al., 1994). Local coherence is established heterarchically, or horizontally, across the representation; global coherence is established hierarchically, or vertically. These two types of coherence are described in more detail in section 3.2.

In the next section, we briefly examine some attempts to describe what coherence is.

3.1. *What is coherence?*

'Coherence' has various meanings, depending on the context where it is used. For example, in diagnosis, the most coherent explanation of a set of data is often the one which is most consistent with regard to known facts (Ng and Mooney, 1990).

In the story comprehension literature, early characterisations of coherence treated it as a property of a story. According to these theories, 'a discourse is coherent because successive utterances are "about" the same entities' (Hobbs, 1979). An example of this characterisation of coherence can be found in the early work of Kintsch and van Dijk (1978). In their comprehension simulation, coherence connections between successive clauses of a text are determined by argument overlap (where 'argument' is taken to mean some discourse entity such as a character or eventuality). Consider the following:

John took a train to Paris. He arrived ten minutes late.

According to the argument overlap criterion, these two sentences are coherent because they are 'about' John. However, as Hobbs (1979) explains, this characterisation of coherence is inadequate. He cites the following example in support of his claim:

John took a train from Paris to Istanbul. He likes spinach.

It is clear that this text is disjointed in some way, and would be regarded as incoherent by most readers. Hobbs views coherence on a continuum, with stories that are difficult to connect being less coherent than texts that encourage and enable connections. Coherence is described in terms of the density and overlap of relations that obtain between pairs of story constituents, as inferred by a comprehender. These include *Elaboration, Contrast*, and *Parallel*, which are common in the literature on coherence relations (Eberle, 1992; Asher, 1993; Dahlgren, 1988).

In general terms, coherence therefore depends on particular relations holding between elements of a story. We examine the dimensions of coherence in the following sections.

3.2. *Local and global coherence*

Researchers frequently distinguish between local and global coherence in the representation of a story. These types of coherence are established by application of knowledge sources with respect to the comprehender's goals. Each type of coherence is described in more detail in the following sections.

3.2.1. *Local coherence*

Local coherence is captured in both linear and heterarchical representations. It 'refers to structures and processes that organise elements, constituents, and referents of adjacent clauses, or short sequences of clauses' (Graesser et al., 1994).

Knowledge structures used to establish local coherence can be split into two groups:

- *Temporal relations*
  These include relations like 'precedes' or 'overlaps' (Dahlgren et al., 1988).

– *Causal relations*
  These include relations like those used by Trabasso and co-workers, e.g. 'physically causes' and 'enables' (see section 2.3.2).

3.2.2. *Global coherence*

Global coherence is hierarchical, and 'is achieved to the extent that most or all of the constituents can be linked together by one or more overarching themes' (Graesser et al., 1994). This level of coherence depends on finding 'higher order chunks' which capture the higher-level connections between already locally-coherent parts of the text (Graesser et al., 1994).

The amount of global coherence that can be established depends on the amount of local coherence that has already been constructed; it 'is a function of the pairwise local coherence' of the low-level statements (Thagard, 1989).

Structures which capture global coherence could be broadly considered as *schemas* (Mandler, 1984; Smith, 1997). Schemas can be roughly divided into *world-based* and *narrative-based: world-based* schemas make statements about pragmatic coherence in the world; *narrative-based* schemas make statements about coherence between elements of a story. Some example schemas are:

– *World-based schemas*
  Examples of world-based schemas include MOPs (Lehnert et al., 1983); scripts (Schank and Abelson, 1978); goals and plans (Schank and Wilensky, 1978); and macrostructures (van Dijk, 1977). (While these schemas are often associated with particular processing methods, our emphasis is instead on shared *content*.)
– *Narrative-based schemas*
  Examples of narrative-based schemas include story grammar categories (Mandler and Johnson, 1977); thematic abstraction units (Lehnert et al., 1983); plot units (Lehnert, 1982); and story points (Wilensky, 1983).

3.3. *Which kinds of coherence are established?*

A question which now arises is to what extent the different types of coherence are established during comprehension. Some researchers hypothesise that only local coherence is normally established during reading (McKoon and Ratcliff, 1992). Others state that readers 'attempt to construct the most global meaning representation that can be managed on the basis of the text and the reader's background knowledge structures' (Graesser et al., 1994).

Singer et al. provide a taxonomy of these models of 'coherence-seeking' (Singer et al., 1994):

– *Strong minimal hypothesis*
  Comprehenders only attempt to maintain local coherence. No attempt is

made to link local text segments into larger structures. (No researchers fully embrace this position, but several tend towards it, e.g. McKoon and Ratcliff, 1992.)

– *Weak minimal hypothesis*
  Comprehenders attempt to locally connect story statements together to maintain coherence. They also seek links between more distant text segments when coherence breaks down locally. This theory denies that comprehenders create globally coherent structures or 'situation models' of what is going on in a text, unless their goals facilitate such constructions (McKoon and Ratcliff, 1992).
– *Constructionist hypothesis*
  Comprehenders make inferences which build a globally-coherent representation of the situation described by the text (Graesser et al., 1994). Comprehenders continually 'fill out' the text, providing connections between local and distant segments. Story grammarians are implicitly employing this hypothesis, with the claim that story constituents are continually being integrated into a global structure (Mandler and Johnson, 1977).

3.4. *How do comprehenders establish coherence?*

Having briefly described the kinds and dimensions of coherence, it remains to describe the process whereby coherence is established. According to van den Broek (1994), 'perception of coherence is a result of a complex problem-solving process in which the reader infers relations among the ideas, events, and states that are described in the text'. The key term here is *inference*: 'The construction of coherence is the result of inferential processes that take place as the reader proceeds through the text' (van den Broek, 1994).

Inferences form connections between story elements, each establishing a tentative or partial connection between one or more elements. For example, the comprehender may determine that two eventualities have a particular 'local' temporal relationship. This temporal relationship could then be used to form a causal connection, according to the principle of temporal priority (see section 2.3.2). Once several causal connections have been established, it may be possible to infer a new 'global' node which subsumes a collection of causally-related eventualities, via a generalisation (see section 2.3.3). The amount and types of inferences made are discussed in the following sections.

4. Story Comprehension and Inferences

What exactly is an inference? Norvig provides one definition:

> . . . any assertion which the reader comes to believe to be true as a result of reading the text, but which was not previously believed by the reader, and was not stated explicitly in the text. (Norvig, 1989)

However, this definition is deficient in two respects. Firstly, it would be useful to know what form the 'assertions' take. A simple answer is that an inference consists of 'arguments and propositions that were not explicitly mentioned in the message' (Singer et al., 1994). Singer shows how inferences are represented using the following story:

Julie soaked the bonfire. It went out.

The explicit nodes representing this story are as follows (adapted from Signer et al., 1994):

(e1, douse, (agt(julie), obj(bonfire)))
(e2, go-out, (exp(bonfire)))

Singer suggests that the following (in italics) are also added to the representation via inference:

(e1, douse, (agt(julie, obj(bonfire), *inst(water)*)))
(e2, go-out, (exp(bonfire)))
(l1, cause(e1, e2))

An instrument has been added to the first event (*water*); and a link, *l1*, between the two events has also been created.

Secondly, not all inferences can be regarded as 'assertions' which the comprehender comes to 'believe'. The extra information provided by inferences may be permanently retained in memory (*encoding*), or only used to temporarily aid comprehension (*activation*). The distinction between activated and encoded inferences is frequently made in the psychological literature (van den Broek, 1994; Kintsch, 1988; Singer et al., 1994). For instance, there is ample evidence that predictive inferences are made, but not incorporated into the representation unless immediately supported by the text (see section 4.3).

Current implementations of story comprehension only infrequently make this distinction. Often, it is embedded in the idea of multiple possible explanations being generated, which are then evaluated. In this case, the explanations which are 'pruned' and not incorporated into the permanent representation could be considered activated, but not encoded.

4.1. *Types of inference*

There are many types of inference which occur during comprehension; Graesser et al. (1994) distinguish 13 different types. Of these, four are directly associated with establishing local and global coherence:

– *Causal antecedent*
  This type of inference finds a causal antecedent for an eventuality, often in the form of a new eventuality. For example, given the story 'The sun was warm. The snowman melted.', a causal antecedent inference would establish that the various constraints on causality hold between the second eventuality and the first, allowing a casual link to be inferred. Causal antecedent inferences contribute to local coherence.
– *Superordinate goal*
  This type of inference establishes a goal that motivates an agent's intentional action. For example, given the text, 'John was hungry. He went into the diner.', a comprehender could infer that John had the goal of eating. Superordinate goals contribute to local and global coherence.
– *Thematic*
  Thematic inferences establish the 'point' or 'moral' of a story. For example, when reading one of Aesop's fables, a comprehender may infer the 'message' of the fable. Thematic inferences contribute to global coherence.
– *Character emotional reaction*
  These inferences establish unstated emotions of characters, caused by eventualities explicitly stated in the story. For example, from the text, 'Mary's balloon burst. She cried and cried.', a comprehender could infer 'Mary was upset'. These inferences contribute to global coherence.

These inference types can be further subdivided, into those which occur in a backwards direction relative to the progression of narrative time (connecting, explanatory inferences), and those which occur in a forwards direction (predictive inferences). We describe both kinds of inference in the next two sections.

4.2. *Backward inferences*

Backward inferences (also called 'explanatory' or 'bridging' in van den Broek, 1990a and Singer et al., 1994) are a way to construct 'reasons why something occurs' (Trabasso and Magliano, 1996). They are central to answering the 'Why?' questions mentioned in section 2.

Backward inferences may simply establish a causal link between the story eventuality currently being attended to and a previously-created node. In other cases, a bridging node may need to be created. (In computational terms, a bridging node is an assumption that is added to the system's database, in order to make an explanation possible.) The addition of bridging nodes to the representation parallels the way marker-passing systems add new nodes that lie on the path between the focal statement and previous coherent chunks of text (see section 5.2.1); and the way abductive systems make assumptions about unstated events (see section 5.3.1).

4.3. *Forward inferences*

There are several reasons why forward inferences should not be made by a comprehender. At first glance, they don't seem to have much purpose because:

- They are redundant. If a comprehender predicts future eventualities, the information supplied by the prediction will often be supplied by the story anyway.
- They are inefficient. It is quite possible that the wrong inference will be drawn.
- They do not contribute to coherence.

Some researchers view forward inferences as being only minimally made, or as not being drawn at all (see Keefe and McDaniel, 1993 for a review). However, other researchers have shown that forward inferences are made, but that they are quickly removed from working memory unless immediately confirmed by the subsequent story (Keefe and McDaniel, 1993). In some cases, they may even establish coherence if a link cannot be made between two eventualities as they stand: seemingly inexplicable eventualities may only be explainable by reference to some expected future eventualities, for example (adapted from Murray et al., 1993):

> The angry waitress was totally fed up with her job. When a rude customer criticised her, she lifted a plate of spaghetti above his head.

These two sentences make more sense with the addition of the causal consequent inference 'She poured the spaghetti over the customer's head'.

4.4. *What is the goal of comprehension?'*

Having looked at the range of possible inferences available from a particular story, it seems as though inferences could be chained together indefinitely. A computational model would soon be swamped with a mass of irrelevant explanations. This doesn't happen with human comprehension: why?

An initial answer is that a comprehender has a *goal* to get from the explicit statements of the story to a representation that maintains coherence, coverage, and important events. Or, in other words, a goal to settle on a subset of the possible explanations space that satisfies their relevance constraints and which respects plausibility considerations (see section 2). In an implemented system, this goal could be an absolute parameter, independently set by the programmer. For example, it may be a requirement to fully instantiate a script (see section 5.1), or to produce a least-cost set of explanations (see section 5.3.1).

An alternative to the single-goal-state approach is to specify a range of acceptable final types of representation, rather than a single type. For example, instead of requiring that a script be instantiated, the comprehender may be able to satisfy their goal by instantiating a script *or* creating a causal network *or* inferring the rhetorical structure of the story (*or* a mixture of all three). This can be implemented by requiring that the representation satisfies one or more of a variety criteria, for example, an 'explanatory quality' threshold (c.f. van den Broek et al., 1995). For example, in a standard problem-solving system, the aim may be to reach an acceptable solution for any given problem, rather than an optimal or complete solution. In contrast to the single-goal-state approach, which requires that a full representation or no representation is constructed, this method places less emphasis on correctness, completeness, or even proof.

This seems closer to human understanding than the strict requirements of current systems. The partial, sufficient-final-state approach could be applied by requiring that the network representing the events in the story is connected sufficiently rather than completely. At this point, the system reaches 'closure' and comprehension is halted. For example, a comprehender might require that any loose ends in the story have been tied up, and that there are no episodes that remain unresolved. On the other hand, they may be content with a very superficial representation, that only contains the 'gist' of the story.

5. Existing Models of Story Comprehension

Now we've reviewed recent evidence from psychology about 'mundane' comprehension, we consider how previous models match up to this experimental evidence. The various models make different statements about comprehension; we examine these statements in terms of how well they answer the questions in the following areas:

- *Representations*
  What counts as a satisfactory representation?
- *Inferences*
  What kinds and amounts of inferences are made? For example, are both forward and backward inferences made? How is the inference process halted?

– *Coherence*
  How is the coherence of the representation measured? How are competing representations evaluated?

5.1. *Script-based models*

Script-based models 'constrain the inferences an understander makes, thereby preventing the process from being swamped in a flood of irrelevancies and redundancies' (Kintsch, 1988). It is the process of constraining which inferences can be actualised, rather than by actualising all possible inferences and then integrating them (see section 5.3.1), that keeps the inferencing process manageable.

The interaction between a story and a script-based comprehender involves the following steps:

1. Begin reading with no pre-conceptions of what is going to be read.
2. Based on triggers in the story, retrieve scripts which could potentially 'cover' and explain the events in the story (so far).
3. Instantiate slots in the script with any available information.
4. Look for information that can fill the remaining unfilled slots.

FRUMP, an example of an implemented script-based model, is described in the next section.

5.1.1. *FRUMP*

FRUMP uses scripts as a framework for skim-reading of news articles about earthquakes and other disasters (DeJong, 1979). The representations produced are equivalent to summaries of the articles read. FRUMP has an *implicit* representation of relationships between eventualities as causal relationships are implicit in the 'chunking' of a group of eventualities under a single script.

A script in FRUMP is 'sketchy', and only contains slots for some items (those considered important by DeJong). Once a script is activated, the predictor module of the system directs the substantiator to look for certain things in the rest of the text (i.e. to find appropriate slot fillers). In this way, FRUMP prioritises certain processing operations over others. When a new sentence is encountered, interpretations are 'coerced' into providing the required information.

5.1.1.1. *Problems with FRUMP*. While this approach is robust, the notion of what is important in the story is buried within the stored structures (in semantic memory). As the stored structures only contain slots that are deemed important for a single task, only one kind of comprehension is possible: behaviour which will fill all or most of the slots in the activated script. It is clear that this strategy only works when the available scripts are 'pre-reduced', with all irrelevant information removed from them. If there were also irrelevant slots in the scripts, the system would attempt to fill these, and this might reduce the efficiency with which it fills the important slots.

Perhaps the most important problem with the script-based approach is its reliance on 'predict-and-substantiate' processing. This goes directly against experimental evidence about inferences: predictive inferences are not generated at point x in the story, unless strongly supported by the story up to point x (Trabasso and Magliano, 1996); in addition, predictive inferences not immediately confirmed by the story after point x are not incorporated into the representation (see section 4.3 and Murray et al., 1993 for more details). While it is difficult to define 'strong support' or 'confirmation', it is clear that an overly-predictive model does not reflect mundane comprehension.

5.2. *Spreading activation*

One of the most influential paradigms for story comprehension is the spreading-activation model, based on Quillian's psychological work on semantic networks (also called associative networks) (Morris, 1978). In AI, such models are called marker-passing systems. We use the broader term 'spreading activation' to draw attention to the fact that some psychological models of comprehension are theoretically similar to implemented marker-passing systems (e.g. Kintsch, 1989).

In marker-passing systems, inferences are made using spreading activation across a semantic network. Spreading activation is not controlled by expectations, but occurs promiscuously wherever there are arcs between nodes. Usually, some limit is placed on the distance a marker can travel across the network (Charniak, 1986; Norvig, 1989).

Inferences (including resolution of anaphora) occur where there are marker collisions at particular nodes, and/or where paths are created between nodes (Alterman, 1985; Charniak, 1986; Norvig, 1989). The results of the spreading activation process are 'pruned' to obtain the most likely interpretation for a story.

There are similarities at the level of competence between spreading activation models and script-based models: both have semantic memory structures which are 'unified' with a story, such that missing information is added by default. The main difference is in the grain size of the semantic representations used for processing. In spreading activation, the unit of processing is a pair of nodes and the links between them; in script-based processing, the unit of processing can be considered a complete network (Minsky, 1975).

NEXUS, an important spreading-activation system, is described in the next section.

5.2.1. *NEXUS*
In this section, we concentrate on NEXUS as an exemplar of the marker-passing paradigm and as an implemented system which demonstrates the important aspects of the spreading-activation model. Although we mainly limit our discussion of NEXUS, the points we make are generally applicable to other similar systems (e.g. Charniak, 1986; Norvig, 1989). (This section is based on the following papers: Alterman, 1985; Alterman and Bookman, 1990; Alterman and Bookman, 1992.)

NEXUS operates on the principle that 'it is possible to collect together events without explicitly working out all the details of their semantic (e.g. causal) relationships'. It collects the events of a story 'in a causally neutral, but causally relevant, form', and 'does not directly represent deep causal features'. This does not mean that the system uses no notions of causality and temporality, just that these notions are quite loose. The representation produced is a network of instantiated nodes, connected by coherence relations which are derived from knowledge about how eventualities and other entities are related to each other in the real world. In this respect, NEXUS concentrates on generating local coherence relations, as defined in section 3.2.1.

Stories are presented to NEXUS as a list of statements in case notation. The representation that has been built so far is stored on a stack, as a list of concept-coherent chunks. NEXUS examines each story statement in turn and tries to connect the eventuality expressed by that statement with previous chunks. This is done by spreading activation across an *Event Concept Coherence (ECC)* Network. This network consists of nodes representing relationships between concepts. The ECC Network is implemented in the program as a set of clauses, for example:

```
(ANTE EAT HAVE ((MATCH AGT AGT) (MATCH OBJ OBJ)))
```

This relation states that an ANTEcedent of EATing something is HAVEing that thing. In addition, the third argument states that an 'antecedent' relationship between eating and having only holds if the agent (AGT) and the object (OBJ) of the eating are the same as the agent and object of the having (respectively).

Each relation a marker passes over is added to that marker's path; this can be considered as activation of an inference. NEXUS can therefore be classified as a promiscuous inference maker, as the only limit on construction of paths is a bound on their length. This bound is enforced by updating the marker's state at each pass, thus recording how far it has spread from its source concept. When the bound is reached, the marker is not allowed to pass to any more nodes. (Note that alternative passing strategies are possible, e.g. Charniak, 1986.)

The inference paths found are evaluated with respect to the constraints on the links in the path. Only one path is incorporated into the episodic representation. This representation therefore consists of nodes and relations 'copied' from semantic memory (Alterman and Bookman, 1992).

5.2.1.1. *Problems with NEXUS*. Spreading activation in NEXUS is limited by an upper bound. In other words, a variant of Occam's razor is being applied: cut off paths that are too long, i.e. discard explanations that include too many premises and/or assumptions (Ng and Mooney, 1990). One problem with this technique is that the correct path may not be found, because it is too long (c.f. heuristic search algorithms).

Even with a limitation on the 'depth' of the network search, a reasonable-sized semantic network could still produce hundreds of possible paths of valid length. Charniak (1986) noted this problem in relation to his WIMP program: 'The problem with marker passing is that it is not obvious if it can do the job of finding important inferences in a very large and interconnected database'. The main effect of this is that the system has to consider 'a lot of garbage', i.e. marginal paths that provide unlikely explanations of a pair of concepts (Charniak, 1986).

So-called 'isa plateaux' also cause considerable problems. These occur where two concepts belong to the same class, for example, 'boy' and 'dog' both belong to the class 'mammal'. Passing markers to find a link between 'boy' and 'dog' will always find a path that goes through 'mammal', which means that this information is value-less most of the time. In more abstract terms, the information attached to the 'mammal' node is too coarse-grained and general to explicate the co-occurrence of 'boy' and 'dog', and is therefore irrelevant for the purposes of establishing coherence.

The use of constraints on the paths in NEXUS helps to remove paths which are unacceptable, but doesn't solve the problem of having too many paths to check. This is a problem if NEXUS (and other marker-passing systems) are to be considered plausible cognitive models. It seems reasonable to assume that some form of strategic control is exercised over which path is taken next, as is done in production systems which employ conflict resolution (Anderson, 1983).

5.3. *Abduction*

Abduction is defined as 'the process of finding the best explanation for a set of observations; i.e. inferring cause from effect' (Ng and Mooney, 1990). More technically, abduction is the process of finding 'a set of assumptions, which, together with background knowledge, logically entails a set of observations' (Ng and Mooney, 1990).

As abduction works from observations to hypotheses, it is comparable to 'deduction in reverse' (Charniak and McDermott, 1985). For example, we might have the following translation of the English sentences, 'John was happy. The exam was easy.' (in Ng and Mooney's notation):

name (j, john).
happy(j).
exam(e).
easy(e).

The task is to find an explanation for John's happiness. We have the following rulebase to help with this:

succeed(X, Y) → *happy(X).*
exam(Y), easy(Y), study(X, Y), take(X, Y) → *succeed(X, Y).*

We can abduce an explanation for John's happiness by assuming that he has been successful at something. so we add *succeed(j, Y)* to the database as an assumption to be proved. We can prove that John (*j*) has been successful at something, because there is an exam in the database (*e*), the exam is easy, and one thing people can be successful at is exams. We do need to make two further assumptions, which are that John actually studied for the exam and took it, i.e. *study(j, e)* and *take(j, e)*. This allows us to prove *succeed(j, Y)* by replacing it with *succeed(j, e)*.

These rules can be considered as an alternative notation for a semantic network, as used in the spreading-activation models considered in section 5.2 (Hobbs et al., 1993).

5.3.1. *TACITUS*

The TACITUS system, developed by Hobbs et al., is the best current example of an implemented system which uses abduction to form explanations of natural language texts (Hobbs et al., 1993). TACITUS is a very 'loose' system, in that it makes no specific claims about the structure of human memory representations, and takes more of an 'engineering' stance towards comprehension. Any means necessary for comprehension are used, whether this means using cognitively plausible mechanisms or useful 'fudges'. (This section is based on the following papers: Hobbs and Kameyama, 1990; Hobbs et al., 1993; Charniak and Goldman, 1989; Norvig and Wilensky, 1990; Stickel, 1989.)

Rules in TACITUS are axioms which encode knowledge of the various strata of language in first order predicate calculus. The rules are flexible enough to express many different kinds of knowledge. Hobbs et al. (1993) give examples of rules that can be used to segment a text, using discourse relations, e.g.:

$$(\forall e_1, e_2) cause(e_2, e_1) \rightarrow Explanation(e_1, e_2). \tag{1}$$

and rules which can derive schemas from collections of statements (our example):

$$\begin{aligned}&(\forall e_1, e_2, e_3, e_4) restaurantvisit(e_1, e_2, e_3, e_4)\\ &\rightarrow enter(e_1) \wedge order(e_2) \wedge eat(e_3) \wedge leave(e_4).\end{aligned} \tag{2}$$

TACITUS derives one or more ‘neutral’ logical forms for each sentence using an ambiguity-preserving parser. The set of logical forms includes all of the possible explanations for a sentence or set of sentences. Many other incorrect explanations are implicit in the neutral forms that are returned, which are pruned from working memory when a single explanation is determined. Like many other adbuctive systems, control of the comprehension process is applied when the logical forms are evaluated.

Each conjunct in the logical form must either be proved from the database, or assumed. Each conjunct is annotated with a cost, which must be paid for the conjunct to be assumed. If a conjunct can be proven by referring to existing world knowledge, this is preferred over assumption. This process of deriving a best explanation via minimisation of assumption cost is called *weighted abduction* (Hobbs et al., 1993).

The final, single representation produced is the logical form which can be assumed and/or proven with the least cost. This representation may include relations between the eventualities express by the text, or even schema-level structure. Theoretically, there is nothing to prevent the representation produced from being any combination of hierarchical and heterarchical structures, providing the axioms are implemented appropriately.

5.3.1.1. *Problems with TACITUS*. As Norvig and Wilensky (199) explain, the cost-based approach actually conflates the cost of assumption and the quality of the explanation. They demonstrate this using the example of a magician pulling a rabbit out of a hat, which has two possible explanations:

1. The rabbit magically appeared in the hat. This explanation perfectly explains the situation, but has a high cost of assumption (i.e. that magic exists).
2. The magician used sleight-of-hand. This has a low explanatory quality, as it doesn’t specify the exact mechanism, but also has a lower assumption cost.

The probability-based approach to abduction fares slightly better, as the probability that magic was used is far less than the probability of sleight-of-hand (Charniak and Goldman, 1989).

However, all of the metrics suggested for evaluation of abduced explanations (cost-based, probability-based, coherence-based) are applied independently of the comprehender's goals, and thus remain inflexible. Intuitively, this is incorrect: 'explanation is more properly viewed as a means for obtaining useful information than as a goal-neutral process' and 'what constitutes the ultimate goodness of an explanation is its ability to satisfy the needs for information that motivated generating it' (Leake, 1994).

The comprehender's goals may completely override criteria like plausibility and cost; or example, Leake cites how far-fetched or obviously-false explanations may be preferred in a humorous context (Leake, 1994). Also, partial explanations may be acceptable in certain contexts. For example, when someone asks for someone else's opinion of a film, they don't necessarily want to know all the low-level causal explanations that made up the film's plot; they would prefer a patchy 'gist'-like explanation which leaves gaps in their representation, but serves their immediate goals. A cognitively-realistic model should therefore be able to change its criteria for a good explanation from 'choose the cheapest explanation' to 'choose the explanation which satisfies a particular goal, even if it requires more costly assumptions than another one'.

6. Evaluation of Current Approaches

There are several interrelated issues which arise from a brief look at existing implementations of story comprehension:

6.1. *Monostratal understanding*

In many comprehension systems, the representation produced is *monostratal*. In other words, evidence of successful comprehension is provided by instantiation of the appropriate type of knowledge structure. For example, in a script-based system, evidence of success is provided by an instantiated script.

This is reasonable from the point of view of being computationally practical, but fails to discriminate between the contributions of different types of information to a representation. A common intuition about our own comprehension of stories is that we didn't grasp the meaning of a particular part of the story, while we understood the rest perfectly well. This implies that at the semantic level, there may be gradations in the quality of the representation, with some parts being well-represented, while others are poorly-represented.

For example, the same story may be represented by a network of causal relations, with no hierarchical structure, i.e. no macrostructures; alternatively, it could be represented by thematic nodes that have been filled haphazardly by numerous assumptions, but which serve well to unify the story as a whole.

6.2. *Perfect recall*

Current systems embed all of the explicit events in the story into their representations. This means that these systems have flawless recall; presumably it also means that every story statement is available for inclusion in a summary. There is evidence that human encoding is not perfect, and that many of the low-level events in the story may be forgotten, not encoded into the episodic representation, or subsumed into higher-level nodes.

6.3. *Inflexibility vs. inefficiency*

Current systems tend to exhibit rigid behaviours. This is related to their inability to comprehend with varying degrees of flexibility. For example, script-based systems tend to be unable to cope with novel texts. If they encounter a text that doesn't fit a stereotyped event sequence, they are unable to fall back on a different strategy. Weaker, rule-based and maker-passing systems suffer from the opposite problem: they have no access to time-saving stored structures that could reduce the amount of information to be considered during each cycle. Every text is treated as if it is new.

Myers et al. have noted that human comprehenders can utilise both of these types of processing: they are able to apply both default, low-level, pattern matching processes, and strategic, high-level processes (Myers et al., 1994). The pattern-matching processes are efficient and fast, but brittle; the strategic processes are slow and inefficient, but powerful. Current systems display one or the other of these types of behaviour: they either apply a pattern, or use high-level rules, but are rarely able to adapt their behaviour to the task at hand.

Associated with this problem is the issue of *incrementality*: at any point during comprehension, two comprehenders with identical knowledge structures may have reached different intermediate understandings. This may be because of the flexibility (or rigidity) of their comprehension processes. For example, one comprehender may hold-off making global, thematic inferences until they are absolutely sure they are warranted; another comprehender may make such inferences promiscuously, and later discover that they are erroneous. Whitney et al. (1991) detail some experimental corollaries of these kinds of processing, in relation to working memory.

7. Requirements for Adaptive Comprehension

Story comprehension models developed in different academic fields are often widely divergent in their cognitive realism and precision. 'Soft' models, often developed by psychologists, are frequently founded on experimentally-supported mechanisms (e.g. spreading activation, semantic networks, schemas). However, these mechanisms often carry many unstated assumptions: for example, the mechanism for establishing coherence may be reduced to 'argument overlap' (as discussed in section 3.1); or causality may be assumed to be a primitive within the system, without an attempt to specify its semantics. On the other hand, 'hard' AI models tend to be much more advanced when it comes to specifying the actual actions carried out by the system, as these actions must be implemented in program code; the abductive systems discussed in section 5.3 are a clear example of this approach. However, AI systems tend to produce monostratal representations and dispense with cognitive realism when it obstructs comprehension.

Psychological research has identified one of the key features of human comprehension: its balance between flexibility and efficiency. Few current AI systems are able to adapt their behaviour to suit changing story types and circumstances. There are also few detailed accounts of how changing the amount the types of processing performed during comprehension can have an impact on the representation produced.

Future models must address these kinds of behaviour, and the systems based on those models must possess the precision of functioning programs while not dispensing with psychological evidence. One principal requirement is that the representations produced by a system are not restricted to a single class of information. A comprehension system must be adaptable, in that its requirements for coherence at the local and global levels may be varied, to capture the goal-driven nature of comprehension. Potential explanations could then be produced and evaluated according to the requirements of the comprehender, and thus contain variable depths of information within each 'semantic class'.

References

Alterman, R. (1985). A Dictionary Based on Concept Coherence. *Artificial Intelligence* **25**(2): 153–186.

Alterman, R. (1991). Understanding and Summarisation. *Artificial Intelligence Review* **5**(4); 239–254.

Alterman, R. & Bookman, L.A. (1990). Some Computational Experiments in Summarisation. *Discourse Processes* **13**(2): 143–174.

Alterman, R. & Bookman, L.A. (1992). Reasoning About a Semantic Memory Encoding of the Connectivity of Events. *Cognitive Science* **16**(2): 205–232.

Anderson, S. & Slator, B.M. (1990). Requiem for a Theory: the 'Story Grammar' Story. *Journal of Experimental and Theoretical Artificial intelligence* **2**(3): 253–275.

Anderson, J. R. (1983). *The Architecture of Cognition*. Harvard University Press: Cambridge, MA.

Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Kluwer Academic: Dordrecht.

Black, J. B. & Wilensky, R. (1979). An Evaluation of Story Grammars. *Cognitive Science* **3**(3): 213–230.

Black, J. B. & Bower, G. H. (1980). Story Understanding as Problem Solving. *Poetics* **9**(1–3): 223–250.

Charniak, E. (1986). A Neat Theory of Marker Passing. *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI-86)*, 584–588. Morgan Kaufmann: Philadelphia.

Charniak, E. & Goldman, R. (1989). A Semantics for Probabilistic Quantifier-Free First-Order Languages, with Particular Application to Story Understanding. *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, 1074–1079. Morgan Kaufmann: Detroit.

Charniak, E. & McDermott, D. (1985). *Introduction to Artificial Intelligence*. Addison-Wesley: Reading, MA.

Correira, A. (1980). Computing Story Trees. *American Journal of Computational Linguistics* **6**(3–4): 135–149.

Crystal, D. (1971). *Linguistics*. Harmondsworth: Penguin.

Dahlgren, K. (1988). *Naïve Semantics for Natural Language Understanding*. Kluwer Academic: Norwell, MA.

Dahlgren, K., McDowell, J. & Stabler, E. P. (1989). Knowledge Representation for Commonsense Reasoning with Text. *Computational Linguistics* **15**(3): 149–170.

DeJong, G. F. (1979). Prediction and Substantiation: A New Approach to Natural Language Processing. *Cognitive Science* **3**(3): 251–273.

Dyer, M. G. (1992). Scripts. In Shapiro, S. C. (ed.) *Encyclopaedia of Artificial Intelligence*, 1443–1460. John Wiley and Sons: New York.

Eberle, K. (1992). On Representing the Temporal Structure of a Natural Language Text. *Proceedings of the 15th International Conference on Computational Linguistics (COLING-92)*, 288-294. Association for Computational Linguistics: Nantes.

Frisch, A. M. & Perlis, D. (1981). A Re-Evaluation of Story Grammars. *Cognitive Science* **5**(1): 76–86.

Garnham, A. (1983). What's Wrong With Story Grammars. *Cognition* **15**: 145–154.

Garnham, A. (1987). *Mental Models as Representations of Discourse and Text*. Ellis-Horwood: Chichester.

Graesser, A. C., Singer, M. & Trabasso, T. (1994). Constructing Inferences During Narrative Text Comprehension. *Psychological Review* **101**(3): 371–395.

Hobbs, J. R. (1979). Coherence and Coreference. *Cognitive Science* **3**(1): 67–90.

Hobbs, J. R. & Kameyama, M. (1990). Translation by Abduction. *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90)*, 155–161. Helsinki University: Helsinki.

Hobbs, J. R., Stickel, M. E., Appelt, D. E. & Martin, P. (1993). Interpretation as Abduction. *Artificial Intelligence* **63**(1–2): 69–142.

Keefe, D. E. & McDaniel, M. A. (1993). The Time Course and Durability of Predictive Inferences. *Journal of Memory and Language* **32**(4): 446–463.

Kintsch, W. (1988). The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model. *Psychological Review* **95**(2): 163–182.

Kintsch, W. & van Dijk, T. A. (1978). Toward a Model of Text Comprehension and Production. *Psychological Review* **85**(5): 363–394.

Leake, D. B. (1994). *Abduction, Experience, and Goals: A Model of Everyday Abductive Explanation*. Technical Report p–95–07, Computer Science Department, Indiana University, Indianapolis.

Lehnert, W. G. (1982). Plot Units: A Narrative Summarisation Strategy. In Lehnert, W. G. and Ringle, M. H. (eds.) *Strategies for Natural Language Processing*, 375–414. Lawrence Erlbaum: Hillsdale, NJ.

Lehnert, W. G., Dyer, M. G., Johnson, P. N., Yang, C. J. & Harley, S. (1983). BORIS – an Experiment in In-Depth Understanding of Narratives. *Artificial Intelligence* **20**(1): 15–62.

Mandler, J. M. (1984). *Stories, Scripts, and Scenes: Aspects of Schema Theory*. Lawrence Erlbaum: Hillsdale, NJ.

Mandler, J. M. & Johnson, N. S. (1977). Remembrance of Things Parsed: Story Structure and Recall. *Cognitive Psychology* **9**(1): 111–151.

McKoon, G. & Ratcliff, R. (1992). Inference during Reading. *Psychological Review* **99**(3): 440–466.

Minsky, M. L. (1975). A Framework for Representing Knowledge. In Winston, P. H. (ed.) *The Psychology of Computer Vision*, 211–277. McGraw-Hill: New York.

Morris, P. (1978). Models of Long-Term Memory. In Gruneberg, M. M. & Morris, P. (eds.) *Aspects of Memory*, 84–103. Methuen: London.

Murray, J. D., Klin, C. M. & Myers, J. L. (1993). Forward Inferences in Narrative Text. *Journal of Memory and Language* **32**(4): 464–473.

Myers, J. L., O'Brien, E. J., Albrecht, J. E. & Mason, R. A. (1994). Maintaining Global Coherence during Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **20**(4): 876–886.

Ng, H. T. & Mooney, R. J. (1990). On the Role of Coherence in Abductive Explanation. *Proceedings of the 8th National Conference on Artificial Intelligence (AAAI-90)*, 337–342. MIT Press: Boston, MA.

Noordman, L. G. M. & Vonk, W. (1992). Reader's Knowledge and the Control of Inferences in Reading. *Language and Cognitive Processes* **7**(3–4): 373–391.

Norvig, P. (1989). Marker Passing as a Weak Method for Text Inferencing. *Cognitive Science* **13**(4): 569–620.

Norvig, P. & Wilensky, R. (1990). A Critical Evaluation of Commensurable Abduction Models for Semantic Interpretation. *Proceedings of the 13th International Conference of Computational Linguistic (COLING-90)*, 224–230. Helsinki University: Helsinki.

Rumelhart, D. E. (1975). Notes on a Schema for Stories. In Bobrow, D. G. and Collins, A. (eds.) *Representation and Understanding*, 211–236. Academic Press: New York.

Schank, R. & Wilensky, R. (1978). A Goal-Directed Production System for Story Understanding. In Waterman, D. A. and Hayes-Roth, F. (eds.) *Pattern-Directed Inference Systems*, 415–430. Academic Press: Orlando, FL.

Schank, R. & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum: Hillsdale, NJ.

Singer, M., Graesser, A. C. & Trabasso, T. (1994). Minimal or Global Inference during Reading. *Journal of Memory and Language* **33**(4): 421–441.

Smith, E. (1997). *A Computational Model of On-Line Story Understanding*. Ph.D. progress report 97–2, School of Computer Science, University of Birmingham.

Stickel, M. (1990). Rationale and Methods for Adbuctive Reasoning in Natural-Language Interpretation. *Proceedings of Natural Language and Logic: International Scientific Symposium*, 233–252. Springer Verlag: Hamburg.

Thagard, P. (1989). Explanatory Coherence. *Behavioral and Brain Science* **12**(3): 435–502.

Thorndyke, P. (1977). Cognitive Structures in Comprehension and Memory of Narrative Discourse. *Cognitive Psychology* **9**(1): 77–110.

Trabasso, T. & Magliano, J. P. (1996). Conscious Understanding during Comprehension. *Discourse Processes* **21**(3): 255–287.

Trabasso, T. & van den Broek, P. (1985). Causal Thinking and the Representation of Narrative Events. *Journal of Memory and Language* **24**(5): 612–630.

Trabasso, T., van den Broek, P. & Suh, S.Y. (1989). Logical Necessity and Transitivity of Causal Relations in Stories. *Discourse Processes* **12**(1): 1–25.

van den Broek, P. (1990a). The Causal Inference Maker: Towards a Process Model of Inference Generation in Text Comprehension. In Balota, D. A., Flores d'Arcais, G. B. and Rayner, K. (eds.) *Comprehension Processes in Reading*, 423–445. Lawrence Erlbaum: Hillsdale, NJ.

van den Broek, P. (1990b). Causal Inferences and the Comprehension of Narrative Texts. *The Psychology of Learning and Motivation 25*: 175–196.

van den Broek, P. (1994). Comprehension and Memory of Narrative Texts: Inferences and Coherence. In Gernsbacher, M. A. (ed.) *Handbook of Psycholinguistics*, 539–588. Academic Press: London.

van den Broek, P., Risden, K. & Husebye-Hartmann, E. (1995). The Role of Readers' Standards for Coherence in the Generation of Inferences during Reading. In Lorch, R. F. and O'Brien, E. J. (eds.) *Sources of Coherence in Reading*, 353–373. Lawrence Erlbaum: Hillsdale, NJ.

van den Broek, P. & Trabasso, T. (1986). Causal Networks versus Goal Hierarchies in Summarising Text. *Discourse Processes* **9**(1): 1–15.

van Dijk, T. A. (1977). Semantic Macro-Structures and Knowledge Frames in Discourse Comprehension. In Just, M. A. and Carpenter, P. A. (eds.) *Cognitive Processes in Comprehension*, 3–32. Lawrence Erlbaum: Hillsdale, NJ.

Whitney, P., Ritchie, B. G. & Clark, M. B. (1991). Working-Memory Capacity and the Use of Elaborative Inferences in Text Comprehension. *Discourse Processes* **14**(2): 133–145.

Wilensky, R. (1983). Story Grammar versus Story Points. *Behavioral and Brain Sciences* **6**(4): 579–623.