Modeling Multiple Levels of Text Representation

Stefan L. Frank
Discourse Studies, Tilburg University, The Netherlands
Center for Language Studies, University of Nijmegen, The Netherlands

Mathieu Koppen
NICI, University of Nijmegen, The Netherlands

Leo G.M. Noordman
Discourse Studies, Tilburg University, The Netherlands

Wietske Vonk
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
Center for Language Studies, University of Nijmegen, The Netherlands

In F. Schmalhofer & C.A. Perfetti (Eds.), *Higher level language processes in the brain: inference and comprehension processes* (pp 133-157). Mahwah, NJ: Erlbaum (2007).

1 Introduction

A broad model of text comprehension should not only simulate how information is extracted from the text itself, but also how this information is interpreted in light of the reader's knowledge. This distinction is related to the distinction among three levels of discourse representation whose existence has been assumed ever since it was proposed by Kintsch and Van Dijk (1978; see also Van Dijk & Kintsch, 1983). The first level is the *surface representation*, consisting of the text's literal wording. This representation gives rise to the second level, called the *textbase*, where the meaning of the text is represented as a network of concepts and propositions from the text (Kintsch, 1988, 1998). Items in this network are generally assumed to be connected to each other if they have some structural feature in common (e.g., two propositions sharing an argument) or if a connective in the text signals that they are connected: "connection relations between propositions in a coherent text base are typically expressed by connectives such as 'and,' 'but,' 'because,' 'although,' 'yet,' 'then,' 'next,' and so on" (Kintsch & Van Dijk, 1978, p. 390).

When textbase elements are combined with elements from the reader's general knowledge, the third level of representation arises. In this *situation model*, relations among items no longer depend on their structural features. While Zwaan (1999) has argued for the importance of perceptual information to the situation model, Kintsch and Van Dijk (1978) claim that relations among items of the situation model (or "facts", as they call them) depend on the effect the items have on one another's probability of occurring: "relations between facts in some possible world ...are typically of a conditional nature, where the conditional relation may range from possibility, compatibility, or enablement via probability to various kinds of necessity" (p. 390).

Several researchers have attempted to show that Kintsch and Van Dijk's three levels are present in the mental representation of discourse (see e.g. Kintsch, Welsch, Schmalhofer, & Zimny, 1990). Compelling evidence comes from a series of experiments by Fletcher and Chrysler (1990). They had subjects read short stories, each describing a linear ordering among five objects. For instance, one of the stories read

> George likes to flaunt his wealth by purchasing rare art treasures. He has a Persian rug worth as much as my car and it's the cheapest thing he owns. Last week he bought a French oil painting for $12,000 and an Indian necklace for $13,500. George says his wife was angry when she found out that the necklace cost more than the carpet. His most expensive "treasures" are a Ming vase and a Greek statue. The statue is the only thing he ever spent more than $50,000 for. It's hard to believe that the statue cost George more than five times what he paid for the beautiful Persian carpet. (Fletcher & Chrysler, 1990, Table 1)

In this example, five art treasures can be ordered by price: rug/carpet, painting, necklace, vase, and statue. After reading ten of such stories, subjects were given from each story one sentence without its final word. Their task was to choose which of two words was the last of the sentence. For the story above, the test sentence was *George says his wife was angry when she found out that the necklace cost more than the* ... and subjects might have to recognize either *carpet* or *rug* as the actual last word of this sentence in the story they read. Since *carpet* and *rug* are synonyms, the difference between them appears at the surface text level only. If subjects score better than chance on this decision, they must have had some kind of mental representation of the surface text.

Alternatively, the choice might be between *carpet* and *painting*. Since these are not synonyms, this comes down to a choice between different propositions: One states the necklace costs more than the carpet, while according to the other the necklace costs more than the painting. Scoring better on this choice than on the choice between *carpet* and *rug* shows the existence of a level of representation beyond the surface text.

In fact, the necklace cost more than both the carpet and the painting. Subjects who erroneously choose *painting* over *carpet* do not violate the situation model since their choice will still result in a statement that is true in the story. However, if the choice is between *carpet* and *vase*, different choices correspond to different situation models. If subjects score better on this choice than on the choice between *carpet* and *painting*, they must have developed a situation-level representation.

Indeed, Fletcher and Chrysler (1990) did find a better than chance score on the choice between synonyms, an even higher score on the choice between propositions, and the highest score on the choice between situation models. This result strongly supports the existence of at least three levels of representation. Nevertheless, most models of discourse comprehension restrict themselves to only one level. For instance, the Resonance model (Myers & O'Brien, 1998) includes only concepts and propositions that originate from the text, and the connections between them are based on argument overlap. No part of the reader's knowledge is included in the model's text representation, so it remains at the textbase level. Other models are concerned with the situation level only. The units of representation in the models by Langston and Trabasso (1999; Langston, Trabasso, & Magliano, 1999) and by Golden and Rumelhart (1993; Golden, Rumelhart, Strickland, & Ting, 1994), which correspond to story events, are connected to each other only if there is a causal relation between the represented events. Since this causal information originates from the reader's knowledge, and not from the text, these models represent texts at a situational level.

There do exist models that combine information from text and the reader's knowledge, and can therefore be said to implement two levels of representation. However, such models are still mainly textbase-oriented. For example, the episodic memory structures computed by the Landscape model (Van den Broek, Risden, Fletcher, & Thurlow, 1996; Van den Broek, Young, Tzeng, & Linderholm, 1999) consist mainly of concepts originating from the text. Inferred concepts can be added, but they need to be given in advance by the modeler. The model cannot explain why or how these knowledge-based concepts are inferred and added to the text representation.

Kintsch's (1988, 1998) Construction-Integration model does include a process for the integration of knowledge items into the textbase representation. However, this model, too, is mainly concerned with the textbase. Kintsch's (1988) referring to the combination of text and knowledge items as the 'enriched textbase' (p. 166) is a case in point. This enriched textbase does contain a few items from the reader's general knowledge, but most of its structure comes directly from text items.

The model by Schmalhofer, McDaniel, and Keefe (2002) is especially noteworthy since it includes all three levels of representation. In contrast to the models mentioned above, its situational level is not a simple extension of the textbase, but a script-based, independently developed representation.

In short, discourse comprehension models, with the exception of Schmalhofer et al.'s, either implement only one level of representation, or model the situational level as identical to the textbase plus only a few added items. In this chapter, we shall work the other way around. First, we present a purely situational model of knowledge-based inferences for story comprehension. In this Distributed Situation Space model (DSS; Frank, Koppen, Noordman, & Vonk, 2003), the representation of a story completely overlaps with the model's representation of knowledge, and is not derived from the textual formulation of the story. Next, it will be shown how surface texts can give rise to the DSS model's situational representations. The textbase, we shall argue, is of less importance, only playing a role in the transformation of text into a situational representation by providing an intermediate representation that is useful for this process.

2 Representing situations: the DSS model

The Distributed Situation Space model simulates how knowledge-based inferences are made during story comprehension. Its main concern is therefore the implementation and use of the reader's world knowledge, and not the representation of a story text. Using world knowledge in a computational model is problematic because the amount of knowledge readers have is simply too large to implement any significant part of. The DSS model avoids this problem by restricting itself to a microworld, all knowledge of which is implemented in the model. Since this microworld is quite restrictive, only very simple stories can take place in it. Nevertheless, it is complex enough to allow for the evaluation of the model's properties.

2.1 The microworld

In our microworld, there exist two story characters, called Bob and Jilly. Table 1 shows the 14 *basic events*[1] describing Bob and Jilly's possible activities and states. Any story taking place in the microworld can be constructed from these events. Note that they do not have a propositional predicate-argument structure. The DSS model is only concerned with units of meaning to which a truth value can be assigned. There is no such thing as a predicate, argument, or concept in the DSS model, because these cannot carry truth values.

Table 1: Fourteen basic microworld events and their intended meanings.

event	meaning
SUN	The sun shines.
RAIN	It rains.
B OUTSIDE	Bob is outside.
J OUTSIDE	Jilly is outside.
SOCCER	Bob and Jilly play soccer.
HIDE-AND-SEEK	Bob and Jilly play hide-and-seek.
B COMPUTER	Bob plays a computer game.
J COMPUTER	Jilly plays a computer game.
B DOG	Bob plays with the dog.
J DOG	Jilly plays with the dog.
B TIRED	Bob is tired.
J TIRED	Jilly is tired.
B WINS	Bob wins.
J WINS	Jilly wins.

Events in the microworld are assumed to follow one another in discrete *story time steps*. At each time step, some events occur and others do not. The combination of all events that occur and all events that do not occur at the same moment in story time is called the *situation* at that time step.

Of course, some situations are more likely to occur than others. For instance, Bob and Jilly are more likely to be outside than inside when the sun shines while the reverse is true during rain. There also exist impossible situations, such as situations in which soccer is played inside or a computer game outside, or both Bob and Jilly win. Also note that soccer and hide-and-seek are always played by Bob and Jilly together (this is why there are no basic events ‘Jilly plays soccer’ or ‘Bob plays hide-and-seek’) while they can play a computer game or play with the dog individually. Apart from constraints among events within a time step, there are constraints on how situations follow each other in story time. For instance, someone who is tired is less likely to win at the following moment in story time.

Microworld knowledge about these regularities within and between situations is not directly implemented. Instead, a realistic sequence of 250 consecutive example situations is constructed, and the world knowledge needed by the model is extracted from this sequence. The representations of basic events follow from their co-occurrence in these example situations, as is explained in more detail below. Contingencies between consecutive situations form the basis for the implementation of microworld knowledge about temporal relations.

2.2 Representing basic events

The DSS model uses distributed representations of events, which means that there is no one-to-one mapping between the represented events and the model’s processing elements. Instead, each event is represented by several elements, and each element forms part of the representation of several events. A mathematically equivalent (and often easier) way to think of distributed representations is to view each representation as a vector in a high-dimensional space. The relations among these vectors mirror the relations among the events they represent.

A well-known example of a distributed representation is Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). In this model, each vector stands for a word, and the distance between two vectors is a measure for the semantic relatedness of the represented words. The high-dimensional space the vectors reside in is therefore called a *semantic space*. In the DSS model, the vector representations stand for microworld situations, and therefore reside in a *situation space*. This is, of course, how the model gets its name.

For the DSS vectors to represent situations, the relations among vectors should reflect relations among the represented situations. But what are situational relations? According to Kintsch and Van Dijk (1978), as quoted in the Introduction to this chapter, facts in the situation model are related by the effect they have on one another's probability. Therefore, a distributed situational representation should consist of vectors that reflect the relations among probabilities of the situations they represent. Below, it is explained how precisely such a representation is developed for use in the DSS model. Also, it is shown how both the conditional and unconditional subjective probabilities of events can be directly computed from their vectors.

Each of the 250 microworld example situations, discussed in Section 2.1, can be denoted by a 14-dimensional binary vector containing a 0 for each basic event of Table 1 that does not occur, and a 1 for each basic event that does. These 250 vectors serve as input to a Self-Organizing Map (SOM; Kohonen, 1995), consisting of hexagonal cells forming a 10×15 grid. Each of the 150 cells is associated with a 14-dimensional weight vector, which is adapted to regularities in the 250 input vectors during an unsupervised training process (see Frank et al., 2003, for details). As a result, each of the 14 values in a cell's weight vector indicates the extent to which the cell belongs to the representation of one of the 14 basic events. Each basic event is thereby represented as a pattern of values, between 0 and 1, over all SOM cells. After training the SOM, the value of element p of the weight vector of cell i, denoted $\mu_i(p)$, is the extent to which cell i is part of the representation of basic event p. Figure 1 shows these so-called membership values for each basic event of our microworld.

We have talked about distributed representations as vectors in a high-dimensional space, but the SOM representations in Figure 1 are two-dimensional areas. However, the two are mathematically equivalent. Each SOM cell can be viewed as one dimension of the 150-dimensional state space $[0,1]^{150}$. A SOM area, defined by membership values $\mu_i(p)$ for all cells i, is thereby equivalent to the vector $\mu(p) = (\mu_1(p), \mu_2(p),\ldots, \mu_{150}(p))$ in the state space.

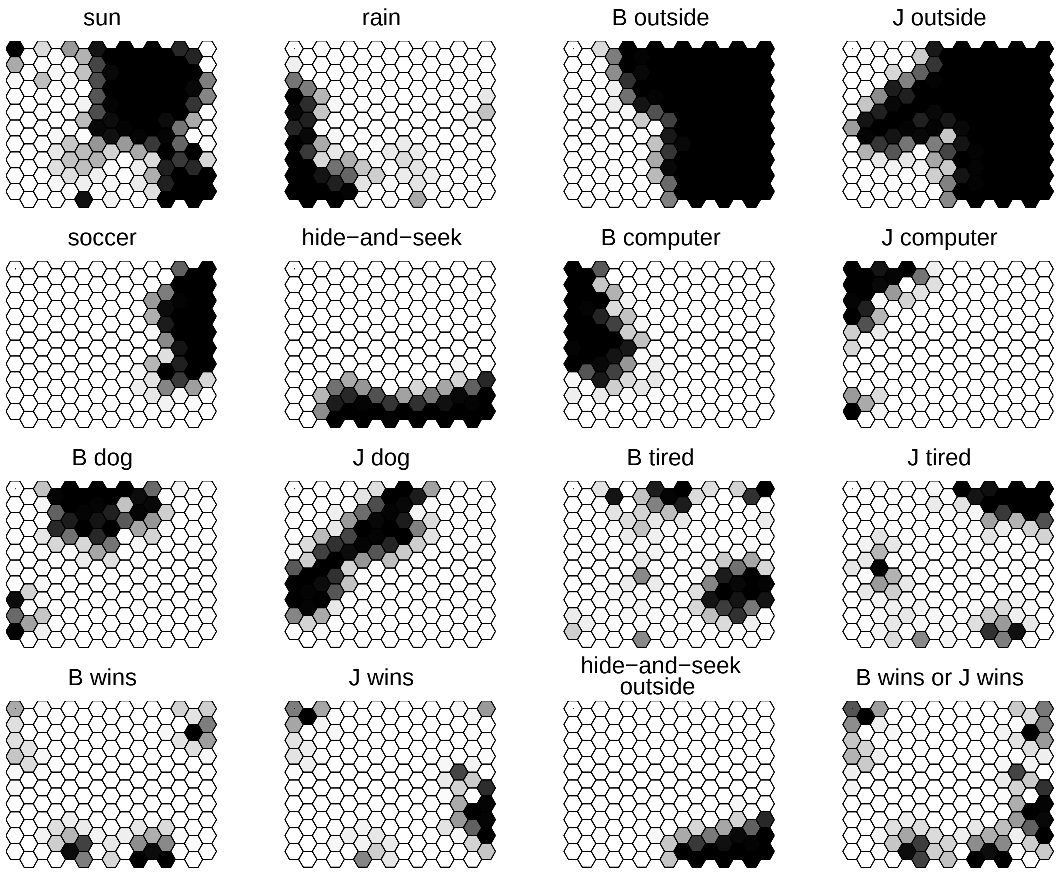

Figure 1: Self-organized mappings of basic events. A cell's membership value for an event is indicated by the cell's darkness. Two examples of complex events are shown in two rightmost mappings of the bottom row: HIDE-AND-SEEK $\wedge$ B OUTSIDE $\wedge$ J OUTSIDE (Bob and Jilly play hide-and-seek outside), and B WINS $\vee$ J WINS (Bob or Jilly wins). Copyright 2003 by the Cognitive Science Society. Reprinted with permission.

2.3 Belief values

The representation of basic events discussed above has some interesting and useful properties. Most importantly, the representations are not arbitrary but closely linked to the probabilities of, and probabilistic relations among, the events. This is what makes them truly situational representations.

Given the vector representation $\mu(p)$ of any event p, it is possible to compute the subjective unconditional probability that p occurs in the microworld. This value, denoted $\tau(p)$, is called the *belief value* of p because it indicates to what extent a reader may belief p to occur at a particular moment in the story. Formally, the belief value of p is computed by

$$\tau(p) = \frac{1}{150}\sum_i \mu_i(p) \tag{1}$$

Also, it is possible to compute the subjective probability that event p occurs in the microworld, given that q occurs at the same moment in story time:

$$\tau(p \mid q) = \frac{\sum_i \mu_i(p)\mu_i(q)}{\sum_i \mu_i(q)} \tag{2}$$

This value is used to determine the meaning of situation vectors. If some situation (vector) X, which does not need to correspond to any basic event, is given, belief values $\tau(p|X)$ for any event p can be computed, giving an indication of what is (not) likely to occur in situation X.

The subjective probabilities (or belief values) correspond very closely to the actual probabilities in the microworld, as Frank et al. (2003) have shown. Using belief values, Frank et al. (2003) define measures of story coherence and of the extent to which an event fits in the story, which are useful for validating the model's results against empirical data.

2.4 Representing complex events

A second important property of the DSS model's representation is that it is productive, in the sense that events can be combined using the Boolean operators of negation, conjunction and disjunction. This means that the representation of any microworld situation can be computed from the representations of basic events. Given a cell's membership values for p and for q, its values for 'not p' and for 'p and q' are computed by:

$$\begin{aligned} \mu_i(\neg p) &= 1 - \mu_i(p) \\ \mu_i(p \wedge q) &= \mu_i(p)\mu_i(q) \end{aligned} \tag{3}$$

Since all connectives in propositional logic can be defined in terms of negation and conjunction, any story situation can be represented using the representations of basic events and the two rules in Equation 3. Figure 1 (bottom right) shows two examples of such complex events.

2.5 Temporal knowledge and the inference process

The DSS model's implementation of temporal knowledge and its inference process are of less importance to this chapter, so we shall discuss these only briefly here. Microworld knowledge about constraints among events *within* a story time step is implemented in the distributed representations of the events sketched above. Knowledge concerning constraints *between* consecutive time steps is also extracted from the sequence of 250 microworld example situations discussed in Section 2.1, and implemented distributively. The model uses this temporal microworld knowledge to infer what is likely (not) to be the case in a story, apart from the facts already mentioned by the story text.

A story is represented as a temporal sequence of story situation vectors, which enter the model one by one. From the model's mathematical basis, it follows precisely how the situation vectors should be transformed to result in the situation sequence that is most likely given temporal microworld knowledge and the constraints put by the story, and thereby increasing the story's coherence (see Frank et al., 2003, for details). During this process, the change in belief values of events can be obtained to ascertain whether they are inferred to occur, or not to occur.

Immediately after a new situation vector enters the model, much can still be inferred, resulting in a high rate of change in the story situations. As model processing time passes, the rate of change in situation vectors decreases until it drops below a threshold level, which is controlled by a depth-of-processing parameter. At that moment, the new story situation is said to be processed sufficiently and the next situation enters the model. The amount of model processing time that was needed to process the situation is taken as a measure of sentence reading time

This implementation is in accordance with the hypothesis that whether or not an inference is made does not directly depend on the type of inference, but on the availability of relevant knowledge, the extent to which the inference contributes to the story's coherence, and the reader's goals. Noordman, Vonk, and Kempff (1992), as well as Vonk and Noordman (1990), present empirical evidence supporting this view.

2.6 Evaluation of the DSS model

Frank et al. (2003) show that the model indeed infers events that can be expected to occur in the story. In the model, the inference process is not driven by a search for coherence but does result in increased story coherence because the story representation is adjusted to be in closer correspondence with world knowledge. Moreover, the model predicts a fair amount of empirical data. It simulates how processing a story situation that is less related to the previous statement results in more inferences and longer processing times, which is in accordance with data by Myers, Shinjo, and Duffy (1987), Sanders and Noordman (2000), and Vonk and Noordman (1990). Increased amounts of inference and processing time also result from increasing the value of the depth-of-processing parameter. This, too, has been found empirically (Noordman et al., 1992; Stewart, Pickering, & Sanford, 2000). The model has been shown easily extendable to simulate several other processes. A model for story retention, based on DSS, simulates how events that contribute least to the story's coherence are the first to be forgotten. Moreover, it correctly predicts that less is recalled as retention time grows, that events are more likely to be recalled if they fit better in the story, that intrusion (i.e., false recall) is more likely for events that fit better in the story, and that these latter two effects increase over retention time. All of these effects have also been found empirically (Bower, Black, & Turner, 1979; Goldman & Varnhagen, 1986; Luftig, 1982; Varnhagen, Morrison, & Everall, 1994).

An extension for pronoun resolution has been shown to simulate how the initial interpretation of an ambiguous pronoun depends on focus, but can be overridden by context information that is inconsistent with the focus (Frank, Koppen, Noordman, & Vonk, 2004). Experimentally, Arnold, Eisenband, Brown-Schmidt, and Trueswell (2000) found a similar time course of pronoun resolution. Moreover, the model can account for empirical data by Leonard, Waters, and Caplan (1997) and Stewart et al. (2000) regarding reading times and error rates, and can explain how these are affected by focus, context informativeness, and depth-of-processing.

The DSS model presents a picture of discourse comprehension that is quite different from the view that has been prevalent since Kintsch and Van Dijk's (1978) influential paper. Like Golden and Rumelhart (1993), we have focused on the situation model and its relation to the reader's knowledge, instead of the propositional textbase and its relation to the text. The model does not deal with propositional structures but explains how general microworld knowledge shapes the interpretation of incoming facts. Consequently, the model lacks any realistic, textual input. The following section makes a beginning at solving this limitation.

3 Representing sentences

The stories processed by the DSS model are represented at a situational level, while the text from which these situations originate is ignored. In this section a first step is made towards extending the model with a more textual level of representation. This shall be accomplished by training a simple recurrent neural network (Elman, 1990) to take as input sentences (i.e., word sequences) describing microworld situations and to transform them into the DSS-vector representations of these situations.

This task is quite similar to the one performed by St. John and McClelland's (1990, 1992) Sentence Gestalt model. One important difference between that model and ours is that the Sentence Gestalt model does not produce a complete representation of the sentence's meaning, but only answers questions about the contents of the sentence. Also, the output representation of the Sentence Gestalt model is localist while ours is distributed.

A model developed by Desai (2002) to simulate language learning by children also consists of a recurrent network that transforms sentences into representations of their meaning. Contrary to the Sentence Gestalt model, these output representations do contain all the information in the sentence. They are, however, still localist. A model even more similar to ours is the Connectionist Sentence Comprehension and Production (CSCP) model by Rohde (2002). It consists of a neural network that, like ours, learns to transform sentences into independently developed, distributed output representations. However, unlike our DSS vectors, the distributed output vectors of the CSCP model were not designed to represent statements at a situational level but only to encode and decode propositional structures. The relations between those vectors do not reflect probabilistic relations between the world events they represent.

The most important respect in which all three of the models mentioned above differ from the one presented in this section is that we shall look at the network's internal representation that develops during training, while the above models are mainly concerned with the training process itself and with the generated output. This internal representation, we shall argue, provides a different view of the traditional surface/textbase/situation-distinction in levels of text representation.

3.1 The microlanguage

The sentences the network learns to process are composed of 15 different "word" units, most of which are words in English: *Bob, Jilly, and, play, be, win, lose, soccer, hide-and-seek, a_computer_game, with_the_dog, outside, inside, tired, awake*. To simplify the already simple language, both *a_computer_game* and *with_the_dog* are considered one word. For further simplification, verbs are not inflected. Note that the microlanguage vocabulary lacks the word *not* and other negations. Kaup and Zwaan (2003) argue that processing a negation involves first constructing the situation model of the corresponding non-negated statement, and then directing attention away from it. Such a two-step process is beyond the network's capabilities.

The 15 words can be combined into sentences following the grammar of Table 2. In total, the microlanguage consists of 328 different sentences. Thirty-eight of these, shown in Table 3, are put aside as a test set. Since the network is not trained on these, it is not shown any sentences in which

- hide-and-seek is played outside (Group 1);
- anyone plays with the dog inside (Group 2);
- *Bob and Jilly* (in this order) play soccer (Group 3);
- *Jilly and Bob* (in this order) play a computer game (Group 4).

Table 2: Grammar of Bob and Jilly's microlanguage.

S	→	NP VP
NP	→	*Bob* | *Jilly* | *Bob and Jilly* | *Jilly and Bob*
VP	→	*play* Game [Place | *and be* State | *and* Result]
	→	*be* Place [*and play* Game | *and* State | *and* Result]
	→	*be* State [*and play* Game | *and* Place | *and* Result]
	→	Result [*and play* Game | Place | *and be* State]
Game	→	*soccer* | *hide-and-seek* | *a_computer_game* | *with_the_dog*
Place	→	*outside* | *inside*
State	→	*tired* | *awake*
Result	→	*win* | *lose*

Moreover, some conjunctions only appear in one of the two possible orders (Group 5). For instance, the network is trained on *Bob play soccer and be tired*, but not on *Bob be tired and play soccer*. Note that the first two groups of test sentences describe situations not mentioned by any of the training sentences, while the last three groups consist of alternative descriptions of situations also present in the training set.

3.2 Training the network

Figure 2 shows the architecture of the recurrent neural network that learns to transform microlanguage sentences into the corresponding microworld situation vectors. The words of a sentence enter the network one by one. Each word is represented locally, by activating one of 15 input units. This activation is fed to the hidden layer, consisting of six units, which also receives its own previous activation state. As a result, the pattern of activation over the six units of the hidden layer forms a representation of the sentence read so far. When the last word of the sentence is processed, the activation pattern over the 150 output units should be the 150-dimensional DSS vector representing the situation described by the sentence, which can be used as input to the DSS model. The activation pattern of the hidden layer after processing a complete sentence shall be called the *intermediate representation* of the sentence, because it lies between the network's word-level input and its situation-level output.

Table 3: Thirty-eight sentences used as a test set.

group	sentence
1	Bob play hide-and-seek outside
	Bob be outside and play hide-and-seek
	Jilly play hide-and-seek outside
	Jilly be outside and play hide-and-seek
	Bob and Jilly play hide-and-seek outside
	Bob and Jilly be outside and play hide-and-seek
	Jilly and Bob play hide-and-seek outside
	Jilly and Bob be outside and play hide-and-seek
2	Bob play with_the_dog inside
	Bob be inside and play with_the_dog
	Jilly play with_the_dog inside
	Jilly be inside and play with_the_dog
	Bob and Jilly play with_the_dog inside
	Bob and Jilly be inside and play with_the_dog
	Jilly and Bob play with_the_dog inside
	Jilly and Bob be inside and play with_the_dog
3	Bob and Jilly play soccer
	Bob and Jilly play soccer outside
	Bob and Jilly play soccer inside
	Bob and Jilly play soccer and be tired
	Bob and Jilly play soccer and be awake
	Bob and Jilly play soccer and win
	Bob and Jilly play soccer and lose
4	Jilly and Bob play a_computer_game
	Jilly and Bob play a_computer_game outside
	Jilly and Bob play a_computer_game inside
	Jilly and Bob play a_computer_game and be tired
	Jilly and Bob play a_computer_game and be awake
	Jilly and Bob play a_computer_game and win
	Jilly and Bob play a_computer_game and lose
5	Bob be tired and play soccer
	Bob be outside and tired
	Bob play hide-and-seek and be awake
	Bob be awake and win
	Jilly play a_computer_game and be tired
	Jilly be tired and inside
	Jilly be awake and play with_the_dog
	Jilly lose and be awake

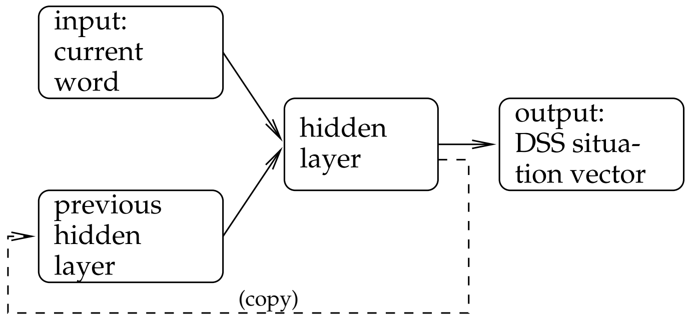

Figure 2: Architecture of the network used to develop intermediate representations of microlanguage sentences. The input layer has 15 units, one for each word. The hidden layer, consisting of six units, receives activation from the input layer and a copy of its own previous state. The output layer has 150 units, one for each situation-space dimension. A solid arrow between two layers indicates that the first layer is fully connected to the second. The dashed arrow indicates that the activations of the hidden layer are copied to the previous hidden layer.

Of course, the network has to be trained to produce the correct situation vector for each input sentence. During training, the output activations are compared to the correct situation vector whenever the complete sentence has been processed. For example, the network is shown the word sequence *Bob play soccer* and produces an output, which is compared to the situation vector $\mu(\text{SOCCER})$.[2] Next, the error in the output is backpropagated to update the connection weights. The set of 290 training sentences was presented to the network 220 times, each time in a different random order. The network is trained seven times, with different random initial weight settings on each occasion. All results presented below are averaged over the results for these seven training sessions.

3.3 Amount of comprehension

To investigate whether the network learned to produce the correct situation vector for each training input, the produced and correct output vectors are compared. This could be done by computing the mean squared error of vector values, but a more easily interpreted measure is available by using belief values. Assume the input sentence describes event p

and the network's output is the vector $X(p)$. If the network has not learned anything, we may expect $\tau(p|X(p))$, the belief value of event p in the situation represented by situation vector $X(p)$, to equal the a priori belief value $\tau(p)$. In that case, the network's 'amount of comprehension' of the sentence is 0. If belief value $\tau(p|X(p))$ is larger than $\tau(p)$, the sentence can be said to be 'understood' to some extent. In the ideal case, when $X(p) = \mu(p)$ so $\tau(p|X(p)) = \tau(p|p)$,[3] the amount of comprehension is defined to equal 1. If, on the other hand, $\tau(p|X(p))$ is smaller than $\tau(p)$, the sentence is misunderstood and the amount of comprehension is negative. Formally, the amount of comprehension of the sentence by the network equals

$$\text{compr}(p) = \frac{\tau(p \mid X(p)) - \tau(p)}{\tau(p \mid p) - \tau(p)} \tag{4}$$

Most microlanguage sentences form a conjunction of two statements. In that case, the comprehension measure of Equation 4 can be somewhat misleading. For instance, if the sentence *Jilly play hide-and-seek outside* results in an output vector that is identical to μ(J OUTSIDE), the network has not understood that Jilly plays hide-and-seek but only that she is outside. Nevertheless, the amount of comprehension will be positive because the belief value of HIDE-AND-SEEK $\wedge$ J OUTSIDE is larger given J OUTSIDE than a priori. Therefore, for sentences describing a conjunction $p \wedge q$, the amount of comprehension is also computed for p and q separately. Note that the amount of comprehension for the individual statements is computed after processing the *complete* sentence, that is, the conjunction $p \wedge q$.

3.4 Results

Learning and generalization

Table 4 shows the average amounts of comprehension for the sentences in the training and test sets, for the complete sentence as well as for the first and second statements separately. All values are significantly positive, indicating that the network does learn to comprehend the training sentences above chance level and generalizes this skill to test sentences. However, first statement comprehension is quite poor, especially for the test sentences. The second statement often seems to overwrite the information in the first.

Table 4: Amounts of comprehension, averaged over n values, and 95% confidence interval for training and test sentences, both for the complete statement and separately for the first and second statement of a sentence describing a conjunction.

		statement		
set	n	complete	first	second
training	2030	.28 ± .01	.18 ± .02	.56 ± .01
test	266	.20 ± .03	.06 ± .04	.62 ± .03

It is also informative to look at the percentages of misunderstood sentences (i.e., resulting in a negative amount of comprehension). The error rates closely follow the amounts of comprehension. Again, first statements are often processed poorly: The error percentages for training sentences are 25.2% and 0.8% for the first and second statement respectively. For test sentences, almost half of the first statements are misunderstood, as can be seen from Table 5. However, these errors are not divided evenly over the 38 test sentences. The network seems to have particular difficulty learning to process sentences that describe new situations (Groups 1 and 2). The first statement of such sentences seems to be completely overwritten by the second. In comparison, the network had more success learning that the connective AND is commutative (Groups 3 to 5), so novel descriptions of previously trained situations are processed reasonably well.

The surprisingly large first statement error rates and negative comprehension scores for Group 2 test sentences can be explained by the microworld situations these sentences refer to. They are all about playing with the dog inside, but Bob and Jilly are more likely to play with their dog *outside*: The a priori belief value of Bob and Jilly being inside is $\tau(\neg(\text{B OUTSIDE})\wedge\neg(\text{J OUTSIDE})) = .26$, while the belief value given that they play with the dog, is only $\tau(\neg(\text{B OUTSIDE})\wedge\neg(\text{J OUTSIDE})|\text{B DOG}\wedge\text{J DOG}) = .12$. This means that understanding only half of a sentence in which Bob and Jilly play with the dog outside will reduce the belief value (and thereby the amount of comprehension) of the other half. Similarly, the large first statement error rates for test sentences in Group 1 are caused by the fact that hide-and-seek is more likely to be played inside, contrary to what these sentences state. Given that they play hide-and-seek, the belief value of Bob and Jilly being inside increases to $\tau(\neg(\text{B OUTSIDE}) \wedge \neg(\text{J OUTSIDE}) | \text{HIDE-AND-SEEK}) = .36$.

Table 5: Error percentages and amounts of comprehension, averaged over *n* values, for test sentences, both for the first and second statement of a sentence describing a conjunction, per test sentence group and averaged over all test sentences. Group numbers refer to Table 3.

		statement			
		first		second	
group	*n*	% errors	compr.	% errors	compr
1	56	64.3	−.15	0.0	.78
2	56	75.0	−.05	0.0	.68
3	49	31.0	.31	0.0	.50
4	49	16.7	.16	2.4	.44
5	56	35.7	.10	1.8	.62
all	266	46.8	.06	0.8	.62

The intermediate representation

Recall from the Introduction the experiment by Fletcher and Chrysler (1990) from which they concluded that there exist three distinct levels of discourse representation: the surface text, the textbase, and the situation model. In this experiment, subjects more often confused two sentences that differed only at the surface-text level than two sentences that differed also at the textbase level. A similar distinction can be made with sentences in our microlanguage. The sentences *Bob and Jilly play soccer* and *Jilly and Bob play soccer* differ at the surface level but, supposedly, not at the propositional level since the commutative property of AND makes AND(BOB,JILLY) the same proposition as AND(JILLY,BOB). Contrary to this, the sentences *Bob play soccer* and *Jilly play soccer* differ both as surface texts and as propositions. They describe the same situation, however, because soccer is always played by both Bob and Jilly.

Eight pairs of sentences about *Bob and Jilly* and their *Jilly and Bob* counterparts form the so-called ‘surface different’ set of sentence pairs, shown in Table 6. The two sentences of each of these pairs differ only at the surface text level. The 10 ‘textbase different’ sentence pairs, also shown in Table 6, describe different propositions but identical situations.

Fletcher and Chrysler’s subjects were also more likely to confuse two sentences that differed only at the surface text and textbase levels than two sentences that differed at the situational level as well. Again, this distinction can be made in our microlanguage. The sentences *Bob play soccer* and *Jilly play soccer*, like those of all other pairs in the ‘textbase different’ set, differ propositionally but not situationally. The sentences *Bob play with_the_dog* and *Jilly play with_the_dog*, on the other hand, differ at both the textbase and the situational level. Ten of such sentence pairs, given in Table 6, form the ‘situation different’ set.

Directly modeling Fletcher and Chrysler’s experiment would require the implementation of some kind of word recognition process. We propose that this difficulty can be circumvented by taking the sentences’ intermediate vector representations and assuming that similar vectors are more difficult to tell apart than dissimilar ones. This implies that similarity in the intermediate representations corresponds to confusability of the sentences.

Table 6: Three sets of sentence pairs. Two sentences of a pair from the ‘surface different’ set differ only in surface text and not propositionally. Sentences of a pair from the ‘textbase different’ set differ propositionally but describe identical situations. Sentences of a pair from the ‘situation different’ set differ both propositionally and situationally.

surface different	
Bob and Jilly play soccer	Jilly and Bob play soccer
Bob and Jilly play soccer outside	Jilly and Bob play soccer outside
Bob and Jilly play hide-and-seek	Jilly and Bob play hide-and-seek
Bob and Jilly play hide-and-seek inside	Jilly and Bob play hide-and-seek inside
Bob and Jilly play a_computer_game	Jilly and Bob play a_computer_game
Bob and Jilly play a_computer_game inside	Jilly and Bob play a_computer_game inside
Bob and Jilly play with_the_dog	Jilly and Bob play with_the_dog
Bob and Jilly play with_the_dog outside	Jilly and Bob play with_the_dog outside
textbase different	
Bob play soccer	Jilly play soccer
Bob play hide-and-seek	Jilly play hide-and-seek
Bob play soccer	Bob play soccer outside
Jilly play soccer	Jilly play soccer outside
Bob play a_computer_game	Bob play a_computer_game inside
Jilly play a_computer_game	Jilly play a_computer_game inside
Bob and Jilly play soccer	Bob and Jilly play soccer outside
Jilly and Bob play soccer	Jilly and Bob play soccer outside
Bob and Jilly play a_computer_game	Bob and Jilly play a_computer_game inside
Jilly and Bob play a_computer_game	Jilly and Bob play a_computer_game inside
situation different	
Bob play a_computer_game	Jilly play a_computer_game
Bob play with_the_dog	Jilly play with_the_dog
Bob play hide-and-seek	Bob play hide-and-seek inside
Jilly play hide-and-seek	Jilly play hide-and-seek inside
Bob play with_the_dog	Bob play with_the_dog outside
Jilly play with_the_dog	Jilly play with_the_dog outside
Bob and Jilly play hide-and-seek	Bob and Jilly play hide-and-seek inside
Jilly and Bob play hide-and-seek	Jilly and Bob play hide-and-seek inside
Bob and Jilly play with_the_dog	Bob and Jilly play with_the_dog outside
Jilly and Bob play with_the_dog	Jilly and Bob play with_the_dog outside

As a measure of dissimilarity of two vectors, the euclidean distance between them is used. For each of the seven trained networks, the distances between the 328 vectors for all microlanguage sentences are normalized to an average of 1. Figure 3 shows the normalized distances between the vector representations of sentence pairs from the three different sets, averaged over seven repetitions of 8 distances for the 'textbase different' set and of 10 distances for the other two sets. The distances follow the percentages of correct responses found by Fletcher and Chrysler: Sentences that differ only in surface text are more similar to one another than sentences that differ also propositionally but not situationally ($t_{124} = 9.38$; $p < .001$), which in turn are more similar to one another than sentences that do differ situationally ($t_{138} = 2.05$; $p < .05$).

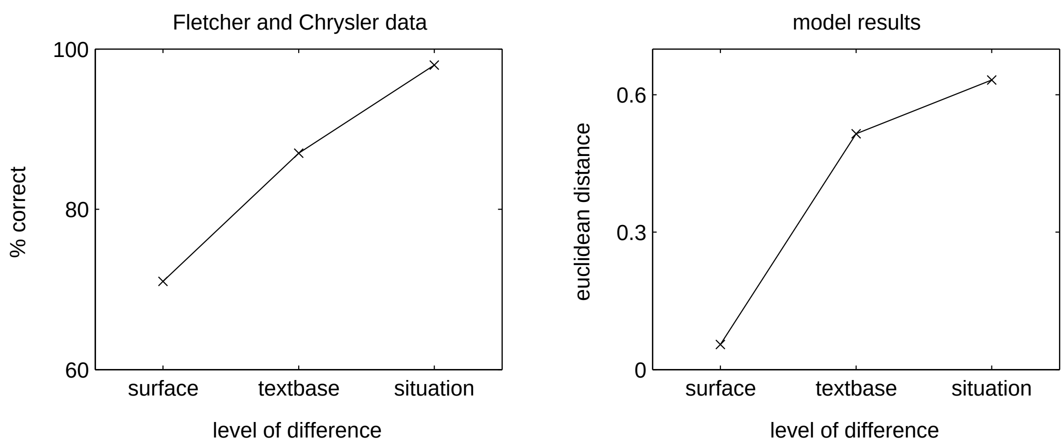

Figure 3: Left: experimental results by Fletcher and Chrysler (1990). Right: distances between vectors representations in the network's hidden layer, for sentences that differ only at the surface level, sentences that differ only at the surface and textbase levels, and sentences that differ also at the situational level.

This effect cannot be explained by a difference in amount of comprehension among the three sets. Although there is an effect of sentence set on the amount of comprehension, this does not follow the same pattern as the effect on distances between intermediate representations. The average amount of comprehension of sentences in the 'textbase different' set (compr = .67) was significantly larger than for both the 'surface different' set (compr = .58; $t_{250} = 3.75$; $p < .001$) and the 'situation different' set (compr = .56; $t_{278} = 4.88$; $p < .001$). The comprehension difference between the 'surface different' and 'situation different' sets was not significant ($t_{250} = .60$; $p > .4$).

3.5 Discussion

Although the network's performance was far from impressive, it did learn to comprehend both training and test sentences significantly above chance level, with the exception of the first statement of sentences describing situations not mentioned in the training set (test sentence Groups 1 and 2). Future research will have to show how generalization performance can be improved. For now, however, we are mainly concerned with the intermediate representation that developed during training.

The intermediate vector representation is neither fully based on surface text nor purely situational. Any difference between two sentences increases the distance between the corresponding vectors, regardless whether the difference is one of surface text, proposition, or situation. Two sentences that describe different situations necessarily also form different propositions, so they will be at least as different from each other as two sentences that differ only propositionally. Likewise, two sentences that differ propositionally must also differ in surface form, so they will be at least as different from each other as two sentences that differ only in surface form. Since any difference between sentences adds to the distance between the corresponding intermediate vector representations, vectors are less similar to one another if the sentences they represent differ at a higher level. This indicates that a *single* representation can encode information about the surface text, the proposition, and the situation. Interestingly, such a representation can predict Fletcher and Chrysler's (1990) findings even though they took their results as providing "strong converging evidence for the psychological reality of van Dijk and Kintsch's (1983) distinction among surface memory, the propositional textbase, and the situation model" (p. 177).

We do not claim that only one level of representation exists. In fact, a purely situational representation was needed to train the network and develop the intermediate representation. It is clear, however, that Fletcher and Chrysler's result does not require three levels of representation to be present at the same time. One property of distributed representations is that they can simultaneously encode different aspects of the represented item. Therefore, a single sentence vector can represent, to some extent, surface text, proposition, and situation.

It is interesting to speculate whether these vectors could account for other experimental data dealing with different levels of text representation. For instance, Schmalhofer and Glavanov (1986) and McDaniel, Schmalhofer, and Keefe (2001) found that reading goal affects the mental representation of text. Unlike the DSS model's story interpretations, the intermediate sentence representations do not depend on the value of a parameter controlling processing depth or reading goal. Each sentence has only one intermediate representation, meaning that the effect of reading goal on this representation cannot be simulated.

Zimny (1987; reported in Kintsch et al., 1990) showed that textual, propositional, and situational information decay from a text's memory trace at different rates. Although this seems to indicate that there are distinct levels of representation, Zimny's results might be accounted for by a single intermediate vector representation. The sentence vector S can be thought of as a location in six-dimensional state space, each dimension of which corresponds to one of the six units in the recurrent network's hidden layer. Surrounding this location is a region all points of which are closer to S than to the representation of any other sentence. We call this the *text region* of S. Also surrounding S is the *proposition region* of S, consisting of all text regions of sentences propositionally identical to S. Since different sentences can correspond to the same proposition, but different propositions cannot be described with a single sentence, all of S's text region must lie within its proposition region. Likewise, S's proposition region is part of a larger *situation region* consisting of the text regions of all sentences describing the same situation as S. Outside this situation region, only sentences describing a situation different from S are represented.

During retention of S in memory, its representation decays. This can be modeled by adding random noise, changing S into a vector we denote S'. More random noise is added as retention time increases, meaning that S' moves randomly through the state space. As long as it stays within *S*'s text region, the literal sentence is remembered. However, sooner or later *S′* will move out of the text region, which means that the surface text of *S* is forgotten. As long as *S′* remains in the proposition region of *S*, however, the proposition described by *S* is known. In other words, the surface text is forgotten more quickly than the textbase. Likewise, when *S′* leaves the proposition region, but is still within the situation region, only the described situation is remembered. This shows that the model may be able to account for other data than Fletcher and Chrysler's (1990).

Of course, all of these findings are tentative. The network only processes single sentences, while a textbase should be able to include several sentences. Also, our microworld and microlanguage are extremely simple, so research is needed with worlds and languages of more realistic size. It is in fact not unlikely that this will increase the textbase-like character of the intermediate representations, since it may be the need to comprehend complex language about relations in a complex world that drives the development of representations that, in some sense, can be regarded propositional. For example, Dennis and Kintsch (this volume) present a system that analyzes natural language texts about a microworld in which different kinds of relations among many entities are of major importance. To answer questions about these texts, the system develops representations that function similarly to propositions, although they do not have predicate-argument structures.

4 Conclusion

The Distributed Situation Space model shows how inferences arise from applying world knowledge to story statements, affecting the interpretation of these statements. In this chapter, we have made a first attempt to extend the model with a simple recurrent network to simulate word-by-word reading. However, much work remains to be done before it can be considered a model for the incremental transformation of word sequences into an interpretation of the story. For instance, the current model cannot predict word reading times, nor does it handle textual processing cues such as connectives and pronouns. Nevertheless, it is able to show how multiple levels of discourse representation can automatically arise from learning the task of transforming a textual representation of a sentence into a situational one.

Traditionally, constructing a textbase is viewed as one of the main goals of text comprehension. Such a textbase is assumed to consist of propositional structures. Our simple recurrent network challenges both these views. First, in our view, understanding a text comes down to constructing a situational representation. Levels of representation prior to the situation model exist only because they are necessary, or at least useful, for transforming a text into a situation model. Second, although the units of meaning at which the text is represented may be proposition-like at the intermediate level, there is no need for propositional predicate-argument structures. A vector that represents a sentence at the hidden layer of our recurrent network does not have such a structure.

Where does this leave empirical evidence of the existence of propositions in the mental representation of a text? Goetz, Anderson, and Schallert (1981) found that subjects often recall all of a proposition or none of it. This has been interpreted as evidence for the cognitive reality of propositions (e.g., Fletcher, 1994; Kintsch, 1998, chap. 3.1; Van Dijk & Kintsch, 1983, chap. 2.2), but it only shows that the units of a text's mental representation may correspond to propositional units. All-or-nothing recall of such propositional units can in fact be interpreted as evidence *against* the existence of propositional *structures*. If subjects never recall part of a proposition, it is very well possible that it does not have any parts. In that case, propositions are represented holistically and not as a collection of related concepts.[4]

Ratcliff and McKoon (1978) performed an experiment designed to show that propositional structures are part of a text's mental representation. They had subjects read sentences such as *The mausoleum that enshrined the tzar overlooked the square*, which consists of two propositions: ENSHRINED(MAUSOLEUM,TZAR) and OVERLOOKED(MAUSOLEUM,SQUARE). If the mental representation of the sentence also contained these propositional structures, so they hypothesized, the words *square* and *mausoleum*, which share a proposition, should prime each other more strongly in a recognition task than the words *square* and *tzar* do, even though the words of this latter pair are closer together in the sentence. Indeed, they did find stronger priming between words that share a proposition than between words that do not, and concluded that propositional structures are cognitively real. However, as is suggested by the above example, they seem not to have taken into account that readers may form a mental image of the events in the text instead of a propositional structure. As noted by Zwaan (1999), the effect on priming might have occurred because, in this mental image, the square and the mausoleum are closer together than the square and the tzar, or even because the tzar, being inside the mausoleum, is not visible from the square.

The same problem occurs in the texts used as experimental stimuli by Dell, McKoon, and Ratcliff (1983). One of these reads

> A burglar surveyed the garage set back from the street. Several milk bottles were piled at the curb. The banker and her husband were on vacation. The criminal slipped away from the streetlamp. (Dell, McKoon, & Ratcliff, 1983, Table 1)

After reading the word *criminal* in the last sentence, recognition of *garage* was found to be faster than after reading a similar text in which *criminal* was replaced by *cat*. This effect was explained by assuming a propositional representation. The text's first sentence gives rise to the proposition SURVEYED(BURGLAR,GARAGE). The anaphor *criminal* in the last sentence refers to the burglar and therefore activates BURGLAR in the reader's mental representation. This results in activation of the concept GARAGE because BURGLAR and GARAGE share the proposition coming from the first sentence.

Such within-proposition activation between concepts can be taken as evidence that the story's mental representation does consist of propositional structures. As in Ratcliff and McKoon's (1978) experiment, however, the stimuli do not seem to have been controlled for the mental image they might evoke in a reader. Experimental findings by Zwaan, Stanfield, and Yaxley (2002) support the hypothesis that the reader of a text constructs a mental image of the scene described by the text. Concepts from the same proposition tend to be close together physically in this scene. In the above example, the burglar is probably very close to the garage in order to survey it. Therefore, focusing attention to the burglar in the mental image of this scene will also highlight the garage.

To conclude, although propositional structures are often assumed, their status in the human cognitive system may not be that well established. Our model shows that a recurrent neural network can learn to transform a string of words into a representation of the story situation described by the words, without needing to first extract any propositional structure from the sentence. The intermediate representation that develops during training the network may be considered a textbase-level representation, but it does not contain propositional structures. Instead, textual, propositional, and situational aspects can be detected simultaneously in this intermediate representation.

Acknowledgements

The research presented here was supported by grant 575-21-007 of the Netherlands Organization for Scientific Research (NWO).

Footnotes

[1] Basic events were called 'basic propositions' in Frank et al.'s (2003) paper.

[2] The network's output to the sentence *Bob play soccer* cannot be compared to the situation vector μ(B SOCCER), because the basic event B SOCCER does not exist. The sentence describes the situation in which both Bob and Jilly play soccer, because they always play this game *together*.

[3] Note that $\tau(p|p)$ can be somewhat less than 1, that is, even a given statement may not be fully believed to be the case.

[4] This does not exclude the possibility that predicates and arguments can somehow be extracted from the intermediate representations, that is, that these are functionally compositional in the sense of Van Gelder (1990).

References

Arnold, J.E., Eisenband, J.G., Brown-Schmidt, S., & Trueswell, J.C. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eyetracking. *Cognition, 76*, B13-B26.

Bower, G.H., Black, J.B., & Turner, T.J. (1979). Scripts in memory for text. *Cognitive Psychology, 11,* 177-220.

Dell, G.S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior, 22*, 121-132.

Desai, R. (2002). *Modeling interaction of syntax and semantics in language acquisition.* Unpublished doctoral dissertation, Indiana University, Bloomington, IN.

Elman, J.L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Fletcher, C.R. (1994). Levels of representation in memory for discourse. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 589-607). San Diego, CA: Academic Press.

Fletcher, C.R., & Chrysler, S.T. (1990). Surface forms, textbases, and situation models: recognition memory for three types of textual information. *Discourse Processes, 13*, 175-190.

Frank, S.L., Koppen, M., Noordman, L.G.M., & Vonk, W. (2003). Modeling knowledge-based inferences in story comprehension. *Cognitive Science, 27,* 875-910.

Frank, S.L., Koppen, M., Noordman, L.G.M., & Vonk, W. (2004). *Context-Sensitive Pronoun Resolution: A Computational Model*. Manuscript submitted for publication.

Goetz, E.T., Anderson, R.C., & Schallert, D.L. (1981). The representation of sentences in memory. *Journal of Verbal Learning and Verbal Behavior, 20,* 369-385.

Golden, R.M., & Rumelhart, D.E. (1993). A parallel distributed processing model of story comprehension and recall. *Discourse Processes, 16*, 203-237.

Golden, R.M., Rumelhart, D.E., Strickland, J., & Ting, A. (1994). Markov random fields for text comprehension. In D.S. Levine & M. Aparicio (Eds.), *Neural networks for knowledge representation and inference* (pp. 283-309). Hillsdale, NJ: Erlbaum.

Goldman, S.R., & Varnhagen, C.K. (1986). Memory for embedded and sequential story structures. *Journal of Memory and Language, 25*, 401-418.

Hinton, G.E., McClelland, J.L., & Rumelhart, D.E. (1986). Distributed representations. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel distributed processing. Volume 1: Foundations* (pp. 77-109). Cambridge, MA: MIT Press.

Kaup, B., & Zwaan, R.A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 439-446.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

Kintsch, W., & Van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Kintsch, W., & Welsch, D.M., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29*, 133-159.

Kohonen, T. (1995). *Self-Organizing Maps*. Berlin: Springer.

Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Langston, M.C., & Trabasso, T. (1999). Modeling causal integration and availability of information during comprehension of narrative texts. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 29-69). Mahwah, NJ: Erlbaum.

Langston, M.C., Trabasso, T., & Magliano, J.P. (1999). A connectionist model of narrative comprehension. In A. Ram & K. Moorman (Eds.), *Understanding language understanding: Computational models of reading* (pp. 181-226). Cambridge, MA: MIT Press.

Leonard, C.L., Waters, G.S., & Caplan, D. (1997). The influence of contextual information on the resolution of ambiguous pronouns by younger and older adults. *Applied Psycholinguistics, 18*, 293-317.

Luftig, R.L. (1982). Effects of paraphrase and schema on intrusions, normalizations, and recall of thematic prose. *Journal of Psycholinguistic Research, 11*, 369-380.

McDaniel, M.A., Schmalhofer, D., & Keefe, D.E. (2001). What is minimal about predictive inferences? *Psychonomic Bulletin & Review, 8*, 840-860.

Myers, J.L., & O'Brien, E.J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*, 131-157.

Myers, J.L., Shinjo, M., & Duffy, S.A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language, 26*, 453-465.

Noordman, L.G.M., Vonk, W., & Kempff, H.J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory and Language, 31*, 573-590.

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior, 17*, 403-417.

Rohde, D.L.T. (2002). *A connectionist model of sentence comprehension and production.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Sanders, T.J.M., & Noordman, L.G.M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes, 29*, 37-60.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: verbatim, propostional, and situational representations. *Journal of Memory and Language*, *25*, 279-294.

Schmalhofer, F., McDaniel, M.A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes, 33*, 105-132.

Stewart, A.J., Pickering, M.J., & Sanford, A.J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language, 42*, 423-443.

St. John , M.F., & McClelland, J.L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence, 46*, 217-257.

St. John , M.F., & McClelland, J.L. (1992). Parallel constraint satisfaction as a comprehension mechanism. In R.G. Reilly & N.E. Sharkey (Eds.), *Connectionist approaches to natural language processing* (pp. 97-136). Hove, UK: Erlbaum.

Van den Broek, P., Risden, K., Fletcher, C.R., & Thurlow, R. (1996). A "landscape" view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B.K. Britton & A.C. Graesser (Eds.), *Models of Understanding Text* (pp. 165-187). Mahwah, NJ: Erlbaum.

Van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71-98). Mahwah, NJ: Erlbaum.

Van Dijk, T.A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Van Gelder, T. (1990). Compositionality: a connectionist variation on a classical theme. *Cognitive Science*, *14*, 355-384.

Varnhagen, C.K., Morrison, F.J., & Everall, R. (1994). Age and schooling effects in story recall and story production. *Developmental Psychology, 30*, 969-979.

Vonk, W., & Noordman, L.G.M. (1990). On the control of inferences in text understanding. In D.A. Balota, G.B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 447-464). Hillsdale, NJ: Erlbaum.

Zimny, S.T. (1987). *Recognition memory for sentences from a discourse*. Unpublished doctoral dissertation. University of Colorado, Boulder, CO.

Zwaan, R.A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes, 28*, 81-88.

Zwaan, R.A., Stanfield, R.A., & Yaxley, R.H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*, 168-171.