The Role of Background Knowledge in Sentence and Discourse Processing
Thesis Proposal*

Raluca Budiu

February 9, 2000

Abstract

One of the challenges of cognitive psychology is developing general models that explain a wide range of empirical phenomena. We propose a unitary theory that addresses data from several text comprehension domains: metaphor understanding, processing of semantic illusions, lexical ambiguity, text memory and elaborative encoding. This theory assumes that sentence processing at the semantic level is incremental, nondeterministic and incomplete, and uses background knowledge hints. While the focus of the theory is on the semantic aspects of language comprehension, we plan to use syntactic ambiguity as a computational constraint. This proposal describes in some detail the sets of data to which this theory has already been applied and the empirical domains that are still to be addressed. We intend to implement this theory as a single model in the ACT-R framework (Anderson & Lebiere, 1998).

1 Introduction

Traditionally, the research on text comprehension has been split in two areas: sentence processing and discourse processing. The term "sentence processing" is used mainly to refer to research on syntactic processing of independent sentences, but it also includes the study of non-literal sentences, lexical ambiguities and of semantic illusions. Discourse (or text) processing addresses questions related to how a coherent representation is built out of a text, i.e. out of a sequence of sentences.

It is commonly agreed that both sentence and discourse processing benefit substantially from background knowledge. One of the principal theories of syntactic processing, the interactive view, states that the probability of attributing a grammatical structure to a sentence depends, among other factors, on the frequency of that construction in the language. Metaphor understanding is deeply rooted in background knowledge: if the grounds of a metaphor like *Time is money* is to be computed, inevitably background knowledge must play a role. Also, background knowledge can lead to errors in understanding. For instance, as Erickson and Mattson found in their study of the Moses illusion (Erickson & Mattson, 1981), people fail to notice distortions in sentences like *How many animals of each kind did Moses take on the ark?* The information given in such a sentence suffices to provide for the retrieval of the ark-story interpretation, and the answer *two* is readily produced.

While the role of background knowledge in discourse processing is widely acknowledged, the literature in the field has been less concerned with how background knowledge exactly contributes to building a text interpretation (with few exceptions — e.g. Sanford & Garrod, 1998). Specifically, the discourse comprehension area has attempted an explanation of how distinct text sentences are put together into a unique text representation (often called situation model or mental model — see Zwaan & Radvansky,

*Additions or major changes to the previous variant of this document are marked with a bar on the left-hand side of the page.

1998 for a review). The situation model is distinct from the traditional schema or script[1] (Schank & Abelson, 1977) — while a schema or a script capture the prototype of a situation, a situation model is the representation of the specific situation encountered in the text. Models of text understanding have focused less on how scripts interfere with the creation of the situation models (but see Sanford & Garrod, 1998 for a counterexample) and more on finding the coherence relations (i.e. inferences, resolving anaphoras etc.) that are needed to construct a complete situation model.

On the other hand, a multitude of studies (Bransford & Johnson, 1972; Bower, Black, & Turner, 1979; Owens, Bower, & Black, 1979; Schustack & Anderson, 1979) have documented the important role that schemas and scripts play in remembering text. This research shows that memory for text can be improved and also deformed by the scripts extant in the background knowledge. One can safely assume that, since the recall of text is influenced by the associated script, that script is accessed when the text is first read and comprehended.

In this thesis we will attempt to address the role that background knowledge plays in sentences and discourse processing. We will claim that background knowledge is involved in similar ways in sentence and discourse processing: namely, the interpretation of a sentence, be it isolated or in discourse, is built incrementally. At each moment, the cognitive system uses all the information available in the sentence to find the right schema or script that identifies the current situation; further on, the content of that schema helps comprehension or recall. Thus, sentence and discourse processing are viewed as a mixture of bottom–up and top–down processes. This model also makes the important claim that building the semantic interpretation of a sentence is powerfully influenced by background knowledge. This view is in contrast with those models of comprehension (Kintsch & Dijk, 1978; McKoon & Ratcliff, 1992) which assume that the integration of the sentence with its context and/or background knowledge is done after the entire sentence has been read and the meanings of the individual words have been processed.

While the thesis will focus on the interaction of language with background knowledge, it would be useful and interesting to explore the ways in which such a theory of comprehension could be extended to account for how text information and background knowledge cooperate in the construction of situation models.

The embodiment of these ideas will be a unique computational theory addressing various phenomena in the area of sentence and text comprehension. Namely, we intend to provide a general model of sentence comprehension and apply that model to empirical data collected from the psychological literature.

In this proposal, we first describe a starting model (formulated for sentence processing); we then continue with the empirical domains to which we hope to apply the model. They can be broadly split into two classes of phenomena: sentence-related and discourse-related. For each empirical field, we briefly present incipient ideas about the extensions that the model may need in order to address data in that field. Finally, we describe in more detail one supplementary computational constraint that the theory needs to satisfy — syntactic ambiguity.

2 The Core of the Model

The starting idea of the thesis consisted of a unified computational model addressing data from the Moses illusion literature and from the metaphor understanding literature. This section describes this simple model, followed in the next sections by the description of its possible applications.

In its initial form, the model is concerned with the comprehension of isolated sentences (no discourse context present). It assumes that understanding an isolated sentence involves finding a structure in memory that matches that sentence. This is equivalent to finding a background knowledge "context" for the current sentence. Namely, when people are asked how many animals of each kind Noah took on the ark, presumably they identify the proper Noah story in their background knowledge repository and give an answer based on that finding. Also, when a metaphor like *Some jobs are jails* is read, people must access some general knowledge pertaining to jobs in order to understand it.

[1] Schank later revised his view on scripts and proposed MOPS which are closer to the situation model (Schank, 1982, 1999).

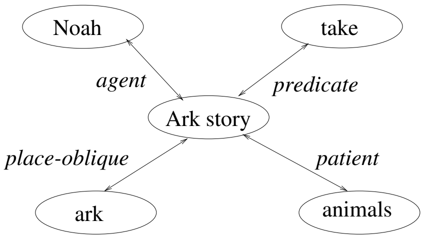

Figure 1: Representation for the background knowledge proposition ("context") *Noah took the animals on the ark.*

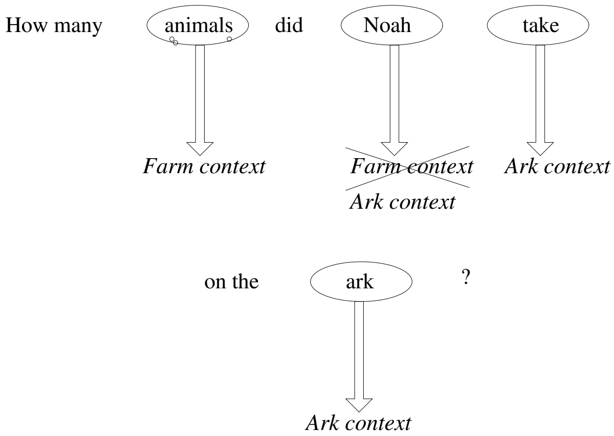

Figure 2: Candidate interpretations for the sentence *How many animals did Noah take on the ark.*

It is certainly true that not all sentences contain information that can be recovered from long term memory. For instance, when we heard the story of Noah for the first time, we did not know how many animals of each kind he took on the ark. New information is certainly added to our knowledge. At this point, we are not concerned with that case; however, we acknowledge that an account of how text representations are formed must deal with it. But for the time being, the model assumes that a sentence is unintelligible if no background knowledge analog can be found for it.

The representation for a background knowledge proposition is exemplified in the Figure 1. The background knowledge fact "Noah took the animals on the ark" is represented as a chunk in the long term memory (the node *Ark story* in the graph). This chunk is connected (via labeled links, which are chunks themselves) to other chunks that represent the concepts involved in this proposition. After a sentence is comprehended, its representation in memory will be similar with the representation of a background knowledge proposition.

The proposed model is incremental. For the moment, suppose that for each word the model knows its correct thematic role, i.e. if it is an agent, a patient etc. With each new word that is read, a candidate interpretation (i.e. a context) for the sentence is sought for. The candidate interpretation must match the actual word meaning and its thematic role in the sentence. Figure 2 shows how the model finds candidate interpretations in the process of reading the sentence *How many animals did Noah take on the ark?* For instance, after reading *animals* a candidate context might be *My father raises ten animals on his farm*[2]. Notice that the current word must have the same thematic role in the new sentence as in the background knowledge proposition that is considered as a candidate context, i.e. in this case *animals* should be a patient in the candidate context, since it is so in the actual sentence[3]. If the next word read is *Noah*, the current candidate context is no longer viable, and a new one, matching *animals*

[2]Many contexts can match at this point. The most active matching context is selected by the model.
[3]The index function of thematic roles will be relaxed further on.

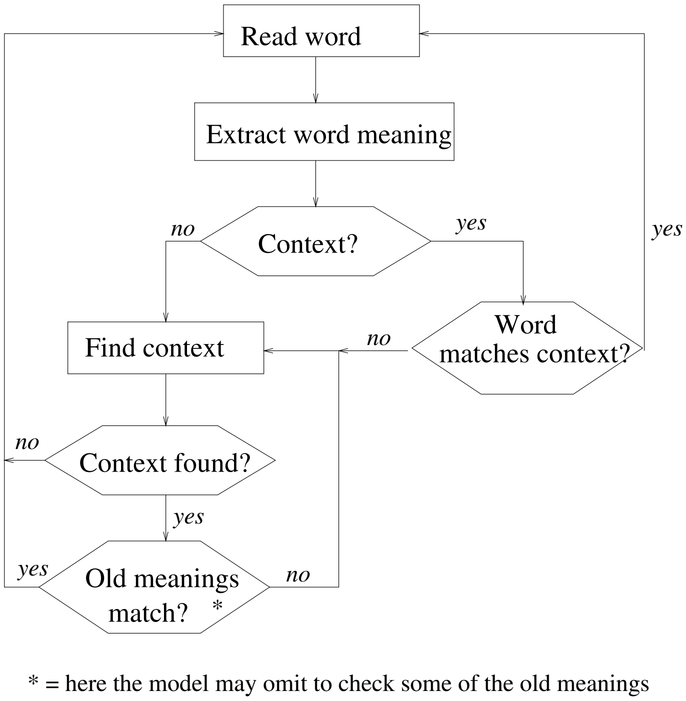

Figure 3: Flowchart of the model in the absence of context priming.

as patient and *Noah* as agent, is sought for. The process repeats until all the words in the sentence have been read.

Figure 3 contains a flowchart describing the behavior of this model. After a new word is read, its meaning is extracted from long term memory. If there is no candidate context, the model looks for a context whose corresponding role matches[4] the current meaning. Coming back to the example in Figure 2, if the current word is *animals* and if its role is patient, the model will search for a context whose patient is a concept similar with *animals*. If no such context is found, the model goes on and reads the next word. If a matching context is found, before accepting it as a candidate, all the concepts from the sentence (corresponding to previous words) are matched against it. If one of them does not match, the current candidate is discarded and another one is sought.

If, immediately after the word is read, there is a candidate context for the sentence, the model checks whether the meaning of the new word matches the corresponding role in the candidate context. Coming back to the example in Figure 2 and assuming that the current word is *Noah*, the model checks whether *Noah* matches the agent of the farm context (i.e. the agent of *My father raises ten animals on his farm*). If the current word matched the current context, the model would go to the next word. However, in this example, *Noah* does not match *father*, so the farm context is abandoned and a new context, matching *Noah* as an agent is looked for. If such a context is found (and it is in the example in Figure 2), the model also checks whether the new context (i.e. the ark story) has a patient which matches the previously processed words (in this case *animals*). If a candidate cannot be found, the next word is read.

If the sentence is long enough, it is possible that the model omits to check all the previously-processed words against a new candidate context[5], and, thus, the context is constrained by only a few of them. The number of words omitted from the check may be an individual-dependent parameter. The essential function of “word-omitting” is to make the comprehension more robust. In the real world, people often

[4] Two concepts match if they are identical or highly similar.

[5] The model executes a loop in which it checks the previous words, but the point of exit out of the loop is chosen nondeterministically.

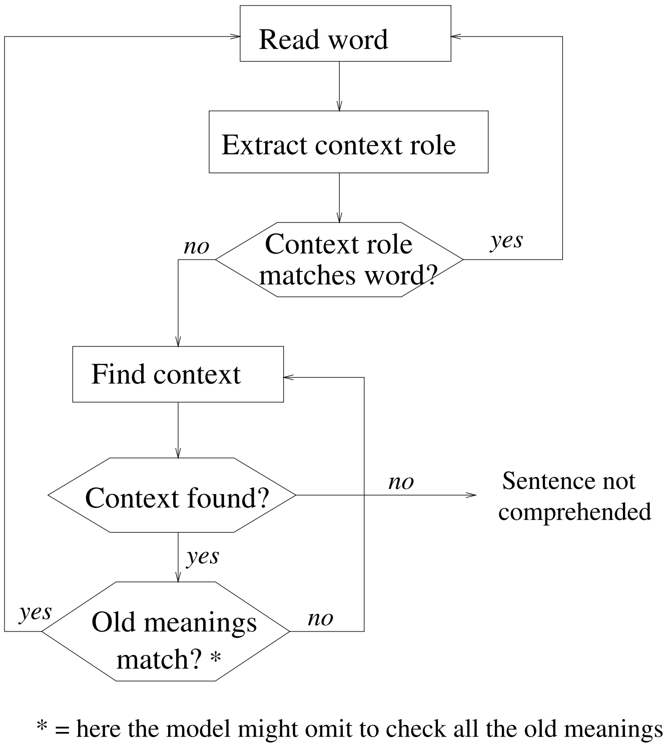

Figure 4: Finding a sentence interpretation with context priming.

ignore some of the words in a sentence (either because they are new words or because of the noise in the environment or because the speaker makes a mistake), and they are still able to find an interpretation for it.

At any point in the sentence, even at the very beginning, the model attempts to find an interpretation for the sentence, using the available information. This type of processing could be called "eager", as opposed to the "lazy" paradigm, in which the sentence representation is first built in its entirety, and then is used to look for a background knowledge antecedent.

Once sufficient words from the sentence have been processed, the model achieves a high degree of confidence that it has the right context, corresponding to the final, correct interpretation. Then the context itself participates in the word comprehension process by priming the meaning of the word considered. The resulting model is described in the flowchart in Figure 4. The main difference between this flowchart and the one in Figure 3 is that, after reading the word, instead of first extracting the word meaning and then matching it onto the candidate context (if any), the corresponding role extant in the context is extracted and matched onto the word meaning. In this way, the context-supported meaning of the word will be always extracted. To see how context priming works, consider again at the example in Figure 2. The Merriam-Webster Dictionary, 1991 mentions that the word *ark* has at least two meanings:

1 a: a boat or ship held to resemble that in which Noah and his family were preserved from the Deluge;

...

2 b: a repository traditionally in or against the wall of a synagogue for the scrolls of the Torah

For the sake of the example, assume that the model is familiar with both meanings of the word *ark*, and that both meanings are equally frequent. Also, suppose that the model has processed all the

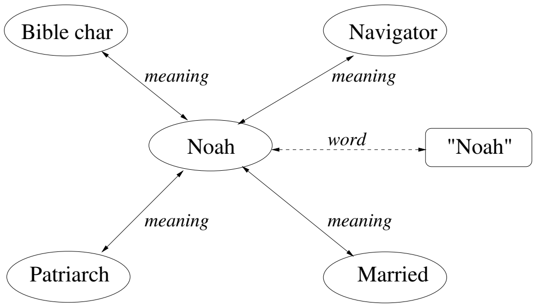

Figure 5: Representation of the meaning of *Noah* as a set of features.

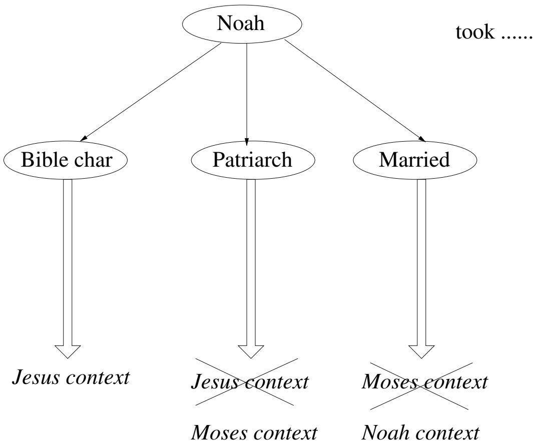

Figure 6: Word features as context matching units when the word *Noah* in the sentence *Noah took two animals on the ark* is processed.

words in the sentence *How many animals did Noah take on the ark* but the last one. At this point, the model has a strong candidate context — the ark story. According to Figure 4, the model extracts the place-oblique corresponding to this context, namely the boat meaning of the word *ark*, and then checks whether this meaning matches the meaning of the word. Thus, by priming one meaning of the word, the context causes the model not to extract the inappropriate meaning.

On the other hand, it is possible that the meaning favored by the context does not match the actual meaning of the word. In that case, the model would need to recalculate the interpretation in such a way as to account for the new word.

If the end of the sentence is reached without the model having found an interpretation for the sentence, that sentence is deemed incomprehensible. However, its representation is added to memory, and can become a context for subsequent sentences.[6]

In an elaboration to this model, we have assumed that the word meaning is a distributed set of features (see Figure 5) and we have made the meaning feature replace the word as the unit for context matching. In this view, processing a word meaning is equivalent with extracting some of its features. The model remains roughly the same; the only change is to replace "word" with "feature" in the flowcharts in Figures 3 and 4. Suppose that the sentence to be processed is *Noah took the animals on the ark*, and that the word *Noah* had the features *patriarch*, *navigator*, *biblical-character*

[6]Apparently, this is nonsensical: how can we understand something in terms of another thing which we did not understand? We agree that the process of assigning an interpretation to a novel fact is more complex than we described here. Thus, even if there is no background knowledge proposition corresponding to the sentence being read, people may be able to interpret it by using analogy, by priming an interpretation via components of the sentence or by more complex mechanisms.

and *married*. Then, as Figure 6 indicates, one might first extract the *biblical-character* feature and consider a candidate context involving a biblical character as an agent (for instance, a proposition about Jesus); after processing the *patriarch* feature, the current candidate context is rejected and a new one, involving an agent who is both a patriarch and a biblical character is looked for (in our example in Figure 2 the new candidate context happens to be the correct one). It's not necessary that all the features of the word *Noah* be actually extracted; the number of features taken into consideration may depend on individual characteristics and on time pressure.

To summarize, the model that we propose has the following qualities:

- is incremental: the interpretation of the sentence is built as the sentence is read;
- follows a trial-and-error paradigm: at each step a candidate interpretation is formed; if subsequently it proves wrong, it is rejected;
- mixes bottom-up with top-down strategies: word meanings help finding a sentence interpretation and the sentence interpretation modulates the meaning processing;
- is based on incomplete retrieval at the word meaning level: the number of word features processed may vary;
- is based on incomplete processing at the sentence level: some of the information previously processed may be actually not taken into account later on and not contribute to the sentence interpretation.

The two last features of the model are in fact instantiations of partial matching: either between the word meaning and the corresponding context role or between the sentence representation and the context representation.

We do not intend to provide a detailed syntactic processing model. One can assume that the syntactic processing is done independently (although not necessarily in a modular way). Thus, according to this view, for each word, syntactic analysis is done first and a syntactical role is attributed to the word. The syntactical information is used in finding a sentence interpretation. However, we must acknowledge the existence of garden path effects: the computation of a syntactical role can produce an incorrect result at a certain point in the sentence, and therefore, the model needs to be extended to allow for such effects.

This theory, as described, has been implemented in the ACT-R framework (Anderson & Lebiere, 1998). ACT-R is a production-system cognitive architecture, which has been widely used to model a variety of problem-solving and memory tasks. The reason for which we chose this architecture is to show that there is nothing special about language processing: it can be implemented in this system in the same way as many other problem-solving tasks. An advantage of the ACT-R theory is that it allows models to produce concrete predictions (reading times, accuracies) that can be compared with the empirical data.

One of the major challenges of modeling sentence comprehension is that people are relatively fast at understanding sentences. The reading time for a typical sentence is on the order of a few seconds. While complicated processes may occur in sentence comprehension, they must happen in this short interval.

The ACT-R implementation of our theory produces sentence processing times comparable to the times that people take in similar tasks. We have tested it against data pertaining to metaphor understanding and semantic illusions, and we plan to extend it to a wider variety of data related to the same phenomena, and also to comprehension of lexical ambiguity, and to text memory phenomena. As an additional computational constraint, we plan for our theory to allow for the comprehension of garden path sentences — that is, to provide a mechanism for error recovery in case of an incorrect thematic role assignment.

Type of sentence	Reading Times (sec)	Predictions (sec)
Metaphor-first	4.21	4.23
Metaphor-second	3.53	2.84

Table 1: Mean reading times and model predictions for metaphorical sentences from Gerrig & Healy (1982).

3 Tests of the Theory: Sentence Processing

3.1 Metaphor Understanding

The literature on metaphor understanding is rich and controversial. A lot of attention has been paid to the question of processing seriality for non-literal sentences. Searle's serial position (Searle, 1979) states that literal comprehension is mandatory and precedes metaphorical comprehension. Only after failing to find a literal interpretation do subjects proceed with searching a metaphorical one.

The psychological literature suggests that often this position is not supported by empirical data. It has been repeatedly shown that in many situations, albeit not all, metaphor comprehension is as fast as literal comprehension. An interesting question in this context is: are there metaphors easy-to-understand and metaphors hard-to-understand[7], and if so, what makes a metaphor harder? Do people select their comprehension strategy according to some perception of metaphor difficulty?

A good starting point in exploring this question is a study by Gerrig and Healy (1983). They found that the position of a metaphor in a sentence affects the reading time of the sentence. Thus, in their experiment, sentences with metaphors at the beginning (like *Drops of molten silver filled the night sky*) took significantly longer to read than sentences in which the metaphors occurred at the end (*The night sky was filled with drops of molten silver*), and this result was not an artifact of the sentence structure. The actual reading times are given in Table 1.

Our model captures this effect. The metaphor-second sentences benefit from the existence of a strong context to guide the meaning processing. Thus, as you can see in Figure 7, before reading *drops of molten silver*, a strong candidate context would be *The night sky was filled with stars*, which expects *stars* to be the theme of this sentence, and therefore enables processing of those features of *drops of molten silver* that are common with *stars*. However, after *Drops of molten silver filled* is read, the candidate context expects a container in the patient role; *sky* does not have many common features with a container, so context priming does not help. Moreover, the current candidate context must be abandoned in favor of a new one, corresponding to the correct interpretation of the sentence. This backtracking takes extra time, and, hence, the model predicts that metaphor-first sentences take longer to comprehend than metaphor-second sentences. As can be seen in Table 1, the ACT-R version of this model correctly reproduces the trends in the data.

We intend to extend the model to capture other data from the metaphor literature. Namely, we would like to be able to model some of the results that contradict Searle's theory. In one study, Glucksberg, Glidea, and Bookin (1982) demonstrated that finding metaphoric interpretations cannot be avoided. They showed subjects sentences like *Some roads are snakes* or *Some flutes are jails* and asked them to judge their literal truth. Glucksberg et al. found that subjects took longer to reject sentences that had a metaphorical interpretation (*Some roads are snakes*) than sentences that had none (*Some flutes are jails*). When asked to recall the sentences they had classified, subjects were better at recalling metaphorically false sentences than other false sentences. The reaction times and the recall rates for these sentences are given in Table 2. Our model can be used to produce the latency result, by assuming that, in order to find the falsity of *Some roads are snakes*, *snakes* needs to be processed more deeply (i.e. to have more of its features extracted) than *jails* in *Some flutes are jails*. The better recall of the metaphors can be explained by the fact that these sentences, unlike the false sentences,

[7] Little attention has been paid to poetic metaphors. Metaphors typically used in psychological studies are ones that could be met in everyday speech.

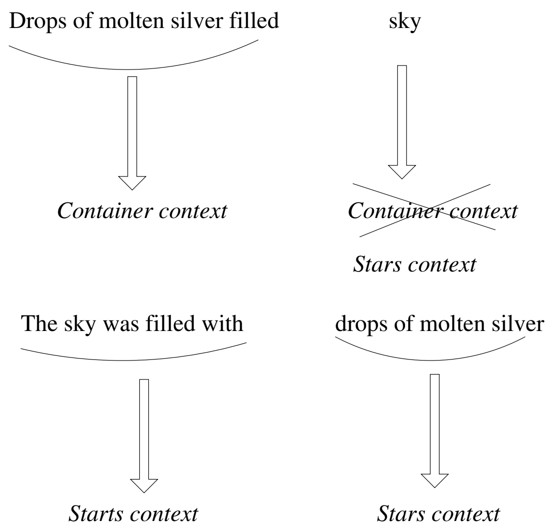

Figure 7: Processing of metaphoric sentences.

Sentence type, with examples	RT(msec)	Recall rate
True (n = 80)	1114	0.249
Some fish are trout		
Some birds are eagles		
False (n = 60)	1174	0.015
Some fish are eagles		
Some jobs are snakes		
Metaphors (n =20)	1239	0.123
Some roads are snakes		
Some jobs are jails		

Table 2: Mean reaction times for making true-false decisions on *Some As are Bs* sentences and recall rates for these sentences (adapted from Glucksberg et al. (1982).

were associated with an interpretation and that interpretation might have served as an extra cue in recall (see also section 5 on text memory for a discussion of similar phenomena).

Another result intended to prove the inevitability of metaphoric processing is reported in Keysar (1989). He manipulated both the metaphorical truth and the literal truth of a sentence by changing the context that preceded it. Subjects were asked to judge the literal truth of the sentences. They showed faster judgement times for congruent sentences (both literally and metaphorically true or both literally and metaphorically false) than for incongruent sentences (see Table 3). Also, subjects were more accurate for the congruent conditions. As before, these results can be explained by noticing that in the incongruent condition there is a greater chance of processing a feature that is non-informative for the correct answer, while in the congruent conditions all features are equally informative. Thus, the longer reaction times for correct responses in the incongruent conditions may be due to the fact that in these cases more features have to be processed in order to get the correct answer. The higher error rates in these conditions also suggest that subjects had a tendency of limiting the number of features processed; it was only in the congruent condition that this tendency did not harm their judgement.

The classic result that contradicts Searle's theory belongs to Ortony, Schallert, Reynolds, and Antos (1978). Their study manipulated the metaphoricity of a sentence by changing the context that preceded it. For instance, the sentence *The hens clucked noisily* could be made metaphorical if preceded by a passage about a women's meeting, or literal, if the preceding passage involved some chicken in a farm.

	Literal Truth	
	True	False
Congruent	874	897
	(.11)	(.06)
Incongruent	954	981
	(.16)	(.16)

Table 3: Mean judgement times (msec) for sentences in Experiment 2 from Keysar (1989). Error rates are shown in parentheses.

	Metaphors	Literals
Long context	2150	1850
Short context	4400	3600

Table 4: Mean comprehension times (msec) for sentences in Ortony et al. (1978). (Times are approximative; the original paper reports data in a graphical form.)

	Reading times (msec)
Metaphor	2,177
Literal	1,735

Table 5: Mean reading times (msec) for sentences in Gibbs (1990).

Ortony et al. found that comprehending literal sentences takes as long as comprehending metaphorical sentences, if the context that precedes each sentence is long enough (see Table 4). Interestingly, if the context was shorter, metaphorical sentences took reliably longer to understand.

However, Gibbs (1990) showed that when the metaphor is anaphoric, (i.e. it is used to refer to a previously introduced character — *The creampuff didn't show up for the match* in a story about a boxing match), there is a comprehension price paid to understand the sentence, possibly because of the extra burden of finding the right topic of the metaphor in the discourse (see Table 5). Questions arise about the exact moment when the metaphor is understood and the reference computed: is it after the noun (*The creampuff*) or after the inconsistent verb (*didn't show up*) or at the end of the sentence. What would happen if the sentence had a literal sense (i.e. if *The hens cluck noisily* was used to refer to a women's meeting, like in Ortony et al., 1978) — would the reference be resolved immediately or postponed? It is probable that the process would be different in this case, since Ortony et al. obtained no difference between literal and metaphorical sentences, in spite of the fact that their metaphors were anaphoric, too. While, for the sake of the theory completeness, we would like to account for these contradictory results, it seems that it would be best done in a discourse processing framework. We plan to address some discourse processing data in the thesis and hope to be able to draw a link from those models back to metaphor processing.

In order to explore further the issue of differences between metaphors, John Anderson and I ran an experiment that compares various types of anaphoric metaphors. In this experiment, we showed subjects passages followed by 4 kinds of sentences. In one type of sentences, metaphoric subject — metaphoric verb sentences, both the subject and the verb were metaphoric, and the sentences made literal sense (e.g. *The hens clucked noisily* after a passage about a women's meeting). This condition is similar with the metaphoric condition in Ortony et al. (1978). Another type of sentences, metaphoric noun + literal verb sentences, was analogous to the metaphoric sentences in Gibbs (1990) (e.g. *The hens talked noisily* in the same context). The other two kinds of sentences were literal subject + metaphoric verb (*The women clucked noisily*) and literal controls (*The women talked noisily*). We collected reading times for the individual words in these sentences. Based on them, we hope to gain microscopic insights

	Illusion Rate (Literal)		%Correct (Gist)	
	Data	Predictions	Data	Predictions
Undistorted	7	8	82	86
Good Distortions	46	50	76	83
Bad Distortions	29	25	74	75

Table 6: Percent correct in the literal and gist task and model predictions for sentences in Ayers et al. (1996).

into the process of comprehension, and thus be able to answer questions referring to the time course of anaphoric metaphor processing. At the macroscopic level, based on the earlier findings, we expect metaphoric subject + literal verb sentences (*The hens talked noisily*) to take longer to understand than literal sentences or than metaphoric subject + metaphoric noun sentences (*The hens clucked noisily*); also, reading times for the latter sentences should be comparable with the reading times for literal sentences. The expectation for the other type of metaphoric sentences (*The women clucked noisily*) is less definite: a study by Gentner and France (1988) showed that verb meanings are more plastic than noun meanings, and therefore, it is possible that verb metaphors will be easier than noun metaphors. Unfortunately, the data from our experiment have not been analyzed yet.

3.2 Semantic Illusions

Erickson and Mattson (1981) were the first to draw the attention on the so-called Moses (or semantic) illusion. Semantic illusions refer to people failing to notice distortions in sentences like *How many animals of each kind did Moses take on the ark?* When confronted with such a question, subjects usually respond *two*, even if they are asked to monitor for distortions and even if they know that it was Noah who took the animals on the ark.

Semantic illusions are a very robust phenomenon: most manipulations intended to make people more aware of the illusion failed (Reder & Kusbit, 1991; Kamas, Reder, & Ayers, 1996). However, not all distortions are hard to notice. When in the ark question, *Moses* is replaced with *Adam*, people are much less likely to fall for the illusion (Ayers, Reder, & Anderson, 1996). Also, certain "illusions" simply don't work: the illusion rate[8] for *Who was the first man to walk on the sun?* is zero.

Ayers et al. (1996) looked at illusion rates for good and bad distortions. They showed subjects three variants of the same distorted question, one containing a good distortion (*How many animals of each kind did Moses take on the ark?*), another containing a bad distortion (*How many animals of each kind did Adam take on the ark?*) and the third being the undistorted question (*How many animals of each kind did Noah take on the ark?*). As expected, subjects fell more often for the good distortions and less often for the bad distortions. They also showed a small bias to call a sentence "distorted" (for undistorted questions, they sometimes gave the "distorted" answer). The illusion rate column in Table 6 presents these results.

As Table 6 shows, the ACT-R model (based on the description in section 2) correctly captures these results. The model assumes that there is a meaning overlap between *Noah* and *Moses*, and also between *Noah* and *Adam*, but the number of salient features that are common to the pair normal – good distortion is greater than the number of salient features common to the pair normal – bad distortion. Figure 8 exemplifies this situation.

The model considers that a sentence is distorted if it is not able to produce an interpretation for it. Thus, when processing a sentence that contains a good distortion (*Moses*), the chance of extracting a feature that does not match the knowledge about the actual ark story is small, since the meaning overlap between the real agent of the ark story (*Noah*) and the distortion (*Moses*) is high. Hence the

[8] The illusion rate is computed as the percentage of times people fell for the illusion, but knew the fact associated to the question.

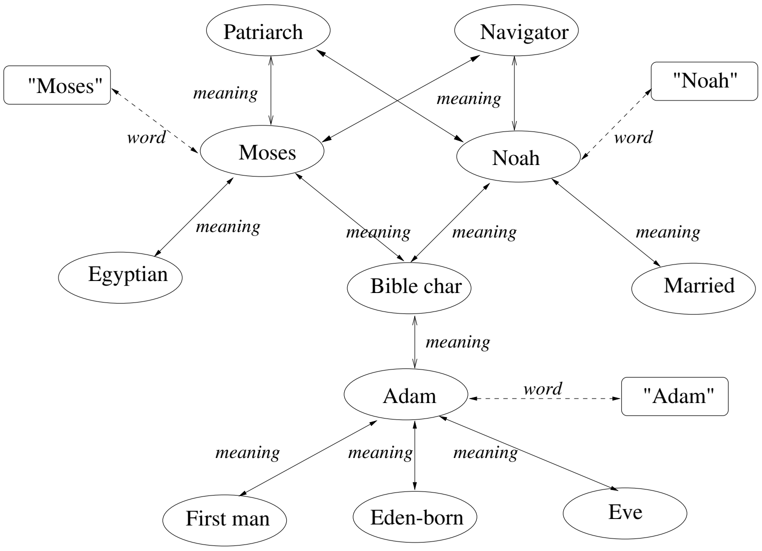

Figure 8: Representation of meanings for *Moses*, *Noah* and *Adam* in the Moses illusion model

model will likely ignore such a feature. Moreover, even if a distinguishing feature is actually processed, it is possible that it be omitted from the check as the model tests a new context candidate.

On the other hand, if the feature overlap between the agent of the ark story and the bad distortion (*Adam*) is small, the chance of stumbling over a distinguishing feature is higher. Also, since these features are in a greater number, the probability of "omitting" all of them at later checks in the subsequent candidate search is small. This model is, to some degree, similar with the explanation offered by Reder (Reder & Kusbit, 1991; Kamas & Reder, 1995; Kamas et al., 1996). Indeed, after examining alternative possibilities, Reder and Kusbit (1991) concluded that the explanation most consistent with their data is a partial matching between the critical concept (i.e. distortion) and the sentential context. The model we propose implements a partial matching mechanism at a symbolic rather than subsymbolic level. Another difference is that our model combines imperfect retrieval (i.e. extracting only the most salient of the features of the word meaning) with partial matching[9]. Kamas and Reder (1995) and Kamas et al. (1996) elaborate the partial matching hypothesis by assuming an activation-spreading mechanism in a semantic network: through that mechanism, "the more activation that accrues at a concept through its connection to the remainder of the concepts in the question, the more likely the person is to accept the retrieved concept as matching the question". Our model does not have an activation-spreading mechanism embedded; the question of enabling it with such a mechanism necessitates further thought.

One interesting prediction that is made by our model is that there is an effect of the position of the distortion on the illusion rate. Indeed, if the distortion appears at the end of the sentence, and if a candidate context exists, that candidate will prime the good, non-distinguishing features of the distortion and the sentence will thus be readily understood. If the distortion occurs at the beginning of the sentence, presumably there is no correct candidate context to guide the feature extraction; hence the chance of processing a distinguishing feature is greater at the beginning than at the end of the sentence.

[9] Notice that we can refer to imperfect retrieval of either the word meaning or the context. Reder and Kusbit (1991) argue against imperfect retrieval of the context based on two experiments in which they manipulated whether subjects learned the correct facts (i.e. *Noah took two animals of each kind on the ark*) before performing the actual Moses illusion task. Those experiments did not increase subjects' sensitivity to distortions. However, this is not necessarily evidence against imperfect retrieval of the word meaning: it is not clear that by this manipulation the actual feature salience distribution for the distortion was changed — i.e. it is possible that, due to comprehension being inherently time-constrained, people tend to actually extract only a few salient meaning features for a word. Experiment 4 in Kamas et al., 1996 is consistent with this hypothesis: if the distinctive features between the distorted and the undistorted terms are made more salient, the sensitivity to the illusion increases.

	Literal	Gist
Normal	4.47	3.33
Distorted	4.49	3.66

Table 7: Mean response latencies (sec) for correct responses in the gist and literal tasks for semantic illusions, taken from Experiment 3 in Reder & Kusbit (1991).

Jaarsveld, Dijkstra, and Hermans (1997) manipulated the position of the distortion in the sentence and did not obtain a significant effect on the illusion rate. They used a verification task, in which Dutch subjects had to judge a sentence like *Moses took two animals of each kind on the ark* or *Two animals of each kind were taken on the ark by Moses.* However, this study did not report baseline percent correct for undistorted sentences, and some hidden effects may have taken place. It is possible that, for instance, subjects had a bias to reject passive sentences (corresponding to final-position distortions) and the potential position effect might have been counterbalanced by this bias. Further investigation of this issue (possibly by collecting new empirical data) is needed.

An interesting variation of the Moses illusion task is the so-called gist task, in which people are asked to ignore the distortions, even if they notice them, and to try to give the "correct" answer. This task is very realistic, in the sense that often times, in the real world, utterances are noisy. Also, in some respects it is similar with the metaphor comprehension task. Unlike for the "literal" Moses illusion task, people are very good at the gist task (Reder & Kusbit, 1991; Ayers et al., 1996); they are also faster to give correct responses in the gist task than in the literal task (Reder & Kusbit, 1991). The column for the percentage of correct answers in the gist task in Table 6 shows some data adapted from the same study by Ayers et al. (1996); latency data from Reder and Kusbit (1991) are shown in Table 7.

It is possible to make the model understand more distorted questions by making it more insistent in looking for an interpretation. Thus, if the end of sentence is reached with no interpretation, instead of deeming the sentence as uninterpretable, the model can continue the search for an interpretation, possibly by dropping features that have been unsuccessful before in helping to find a context. The model is actually very good at performing the gist task; in order to capture the less than perfect performance of subjects we had to assume a small bias of giving the "distorted" answer (see the last column of Table 6 for predictions).

Related to the Moses illusion is the task devised by Barton and Sanford (1993). They asked subjects *When an aircraft crashes, where should the survivors be buried?*; about 77% of the subjects failed to discover the anomaly. However, when the scenario was changed to a bicycle accident (*When a bicycle accident occurs, where should the survivors be buried?*), only 20% fell for the anomaly. Sanford and Garrod (1998) explained these data as possibly due to a role confusion: in the aircraft scenario, there is a *dead* role and a *survivors* role, which can be confused; however, the bicycle accident scenario has no *dead* role, so it makes it easier to detect the anomaly.

In order to account for these data with our model, one could assume that, like in the case of bad distortions for the Moses illusion, there is no background knowledge proposition that relates a bicycle accident with a burial. For the aircrash scenario, however, there are such propositions; moreover, there is a degree of similarity between "dead" and "survivors" (one being the negation of the other). Hence, it can be harder to find an interpretation for the bicycle question than for the aircrash question.

While the model successfully simulates the behavior of the subjects in the semantic illusion task, it makes one fundamental assumption: that what distinguishes a good distortion (*Moses*) of a bad distortion (*Adam*) is that the former has more features in common with the undistorted term (*Noah*). This claim is actually unnecessarily strong: since the model never matches directly two separate meanings (like *Moses* and *Noah*), but only one meaning against a role in the context (i.e. *Moses* against the agent of the ark story, as it is represented in the background knowledge), it is sufficient to assume that the overlap between the distortion and the corresponding context role is greater for good distortions than for bad distortions. Oostendorp and Mul (1990), Oostendorp and Kok (1990) show that such a presumption is a realistic one. In these studies, they collected ratings of similarity between the distortion and the context in which it was supposed to appear (i.e. between *Moses* and the ark story), and they showed that subjects were more likely to fall for those distortions with high context similarity.

4 Tests of the Theory: Lexical Ambiguity

The context-priming feature of our model (see Figure 4) throws it in the disputed battle field of ambiguous word processing. Two views of lexical access have been proposed: the multiple access view maintains that both meanings of an ambiguous word are activated, while the selective access view proposes that only the context-supported meaning is activated. The evidence is split: initially, a number of studies (Swinney, 1979; Onifer & Swinney, 1981) reported data favoring the multiple access view, while more recent research (Tabossi, 1988; Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989) indicates that the time course of ambiguity resolution may be influenced by prior context. The literature is consistent with a model in which both meanings of the ambiguous word are activated, but the amount of activation depends additively on both the individual meaning frequency and on the strength of the preceding context, in spite of the difference of activation not reaching statistical significance in some studies (see Simpson, 1994 for a review of the literature and Lucas, 1999 for a meta-analysis of the various studies published on this topic). As we hope to show by the end of this section, such a view can be accommodated by our model.

Lexical ambiguity studies have been carried using several experimental paradigms. Most of them used a cross-modal priming procedure in which participants had to make a lexical decision or to name a word while they were listening to a sentence containing a homophone. This task resembles a dual task, and therefore, it is possible that processes that are not comprehension-specific be involved (see Kellas, Paul, Martin, & Simpson, 1991 for a discussion). For this reason we will not attempt to model such studies, but we will rather focus on reading latencies and eye-movement data. The assumption with eye movement studies is that longer reading latencies for ambiguous words reflect multiple access.

A number of experiments conducted in Rayner's laboratory (Duffy et al., 1988; Rayner & Frazier, 1989; Rayner & Duffy, 1986) show that the context can influence the meaning selection for an ambiguous word. These studies overlap to some degree, so we only discuss here the most complex one: Duffy et al. (1988). Duffy et al. manipulated the relative frequency of the two meanings of a homophone (balanced versus unbalanced) and the position of the disambiguation (before or after the homophone) in the sentence. They also matched (in terms of frequency and length) each ambiguous word with a control. Some example materials from Duffy et al. (1988) are shown in Table 8. Notice that, for unbalanced words, the sentences always used the subordinate meaning.

Table 9 shows the mean gaze duration on homophones, and Table 10 contains the mean time per character spent on the disambiguating region. To summarize these results:

- when the disambiguation preceded the homophone:
  - there was no difference between gaze duration on balanced homophones versus controls, suggesting that only the correct, context-supported meaning was accessed;
  - subjects gazed longer on unbalanced homophones than on controls, indicating that both meanings were actually accessed;

  There was no difference in the time spent in the disambiguation region.

- when the disambiguation followed the homophone:
  - the gazes on the targets were longer for balanced homophones than for controls; also subjects spent more time in the disambiguation region for sentences containing homophones. These data indicate that both homophone meanings were accessed, and that many times the wrong meaning was selected.

	Balanced homophones:
Disambiguation Before:	Because it was kept on the back of a high shelf, the *pitcher*(whiskey) was often forgotten.
Disambiguation After:	Of course the *pitcher* (whiskey) was often forgotten because it was kept on the back of a high shelf.
	Unbalanced homophones:
Disambiguation Before:	When she finally served it to her guests, the *port* (soup) was a great success.
Disambiguation After:	Last night the *port* (soup) was a great success, when she finally served it to her guests.

Note. The ambiguous target word is italicized. The corresponding control word is included in parentheses.

Table 8: Example sentences used in Duffy et al.(1988).

	Position of disambiguation			
	Before		After	
	Ambiguous	Control	Ambiguous	Control
Balanced	264	264	279	261
Unbalanced	276	255	261	259

Table 9: Mean gaze duration (in msec) on homophones in Duffy et al. (1988).

	Position of disambiguation			
	Before		After	
	Ambiguous	Control	Ambiguous	Control
Balanced	36	36	57	50
Unbalanced	37	36	68	54

Table 10: Mean time (in msec per character) spent on disambiguating region in Duffy et al. (1988).

- subjects spent comparable amounts of time gazing on unbalanced homophones and controls, which suggests that they actually extracted only the dominant meaning of the ambiguous word. Indeed, for unbalanced homophones, subjects spent more time in the disambiguating region, indicating that they had made a wrong sense selection from which they needed to recover.

This pattern of results argues for a view in which word meanings have activations proportional with their frequencies, and the context adds up to the activation of a particular meaning, making it more salient. Our model can accommodate this view by assuming that each meaning of a homophone has an individual activation which is proportional with its frequency. An additional assumption that the model needs to make is that when two chunks have comparable activations and they are both candidates for a retrieval, there is a competition among them that slows down the retrieval process. This last assumption is not one made by the current ACT-R 4.0, but has been proposed as a possible and plausible modification. Given these premises, the case when the disambiguation follows the homophone can be accounted for by using the diagram in Figure 3. In this diagram, since the context is not strong enough, there is no context priming and therefore it is the word meaning that leads the interpretation process. Thus, if a word is a balanced homophone and it precedes the disambiguation, there will be close competition between its two meanings and the model will spend more time processing that word. Moreover, half of the time the model will choose the wrong meaning and therefore will need to recover from that mistake later on when the disambiguation will be processed[10]. Due to this process of recovery which potentially overlaps with the processing of the disambiguating region, the model will spend more time on the disambiguation. For the unbalanced homophones, the wrong meaning will be always selected (as it is the more active) in the disambiguation-after condition and again that will reflect in longer times spent for processing the disambiguation.

However, if the disambiguation precedes the homophone, the flowchart in Figure 4 applies. In this case, the context will guide the meaning retrieval process and will actually be an extra source of activation for the correct word meaning. If the correct word meaning becomes thus the most active (i.e. in the case of balanced homophones), there will be no time penalty for the processing of the homophone. If, however, the activations of the two meanings become comparable (i.e. in the case of unbalanced homophones), longer time will be needed for the actual retrieval of the word meaning.

5 Tests of The Theory: Text Memory

5.1 Script effects

Since Bartlett (1932) a lot of studies have shown the influence of prior schemas on text memory. For instance, Bransford and Johnson (1972) showed that memory for text is improved if a setting is explicitly mentioned, allowing the readers to use their old knowledge about the topic in order to understand the text. Also, Owens et al. (1979) found that when subjects were shown a series of distinct episodes (e.g. in the kitchen, at the doctor's, at supermarket etc.) linked together by a setting mentioned before the first episode (e.g. referring to a pregnant college student), the setting modulated the recall. Namely, not only the recall improved in the presence of a setting, but also the number of setting-specific intrusions made by subjects increased (see Table 11 for the actual data) . The authors also report that, independent on the presence of the setting, subjects made episode-specific intrusions, i.e. whether they knew that the main character was pregnant or not, they recalled propositions about typical supermarket actions that were not studied (e.g. they recalled about the character paying at the check-out counter, even if this was not mentioned). However, they tended to recall better those propositions that could be related with the character's motivations (for instance, they remembered the nurse examination in the doctor episode, as this fact seemed to be more relevant for the pregnancy scenario); also, subjects deformed the propositions studied to be consistent with the character's motive (i.e. they recalled that the nurse ran some *pregnancy tests* instead of *usual procedures*). Also, the recall of the individual episodes was affected by how relevant each episode was for the overall setting (e.g. subjects recalled better the doctor episode, which played an important role in the pregnancy setting). Owens et al. found similar results for a recognition task. This experiment throws some light on the possible interactions between a larger script (the setting) and its subscripts (the individual episodes).

In a related study, Bower et al. (1979) showed that it is possible to increase the number of script[11]-specific intrusions by showing subjects several stories that use the same script (e.g. a story about a visit to a doctor and another about a visit to a dentist). Moreover, as Bower et al. put it, "having another version of the same script mention an action increases the probability that the unmentioned analogous action is intruded in recalling a related story." Data from their experiment is shown in Table 12. Interestingly, the relation between the number of script versions seen and the intrusions from other scripts is not linear (see rows 2 and 3 from Table 12): showing two or three versions of the same script did not produce a significantly different number of intrusions from other scripts. Bower et al. reported similar results for a recognition task (see Table 13) — namely, subjects were more confident

[10] This recovery process will need a more detailed specification in the final model, since in the current variant, the meaning of a word, once retrieved, is not reconsidered.

[11] Unless otherwise noted, we will assume that "schema" and "script" have the same meaning: namely, a prototypical situation stored in long term memory.

	Recalled			
	Nr of episodes	Text Propositions	Intrusions	Total
Before cue				
Setting	3.67	29.24	15.20	44.44
No setting	2.50	20.24	3.76	24.00
After cue				
Setting	4.56	38.44	17.56	56.00
No-setting	4.75	54.63	8.88	63.51

Note. The cue refers to subjects being given episode titles (e.g. *At the supermarket*). The total number of episodes was 5. The total number of propositions was 190.

Table 11: Recall measures in Owens et. al (1979).

Number of script versions	Data		Predictions	
	Stated actions	Unstated actions	Stated actions	Unstated actions
1	.38 (3.03)	.07 (.8)	.35 (2.80)	.05 (.60)
2	.28 (2.27)	.11 (1.26)	.37 (2.96)	.09 (1.08)
3	.32 (2.56)	.10(1.16)		

Note. Number of script versions refers to the number of different stories using the same script.

Table 12: Rate of recall per script version adapted from Bower et al. (1979) and model predictions. Number of actions is shown in parentheses.

	Stated script actions	Unstated script actions	Other actions
1 script version	5.46	3.91	1.71
2 script versions	5.40	4.62	1.76
3 script versions	5.59	4.81	1.86

Note. Number of script versions refers to the number of different stories using the same script.

Table 13: Average recognition ratings (on a scale from 1 to 7) per script version from Bower et al. (1979). 1 means "Very sure I did not read this sentence"; 7 means "Very sure I did read it."

that they had studied script-consistent foils when they were exposed to more than one script versions. In the recognition task, there were three types of foils for which the subjects fell: propositions that mixed up actors and objects from the two stories (e.g. what was true of one character of one story was stated about another character from the second story), propositions not present in any of the stories, but present in the script, and, most interestingly, abstract analogs of propositions present in the other version of the story (e.g. *The nurse checked John's blood pressure and weight* as an analog of *The dental hygienist checked and x-rayed Bill's teeth*).

The ideas in the sentence processing model can be combined with the explanation in Bower et al. (1979) to produce a model of script recall. In this situation, background knowledge corresponds to a set of schemas or scripts. A schema is made of a set of propositions that are related. Understanding a text means finding the right schema to which it corresponds in the background memory. The schema plays the role of the context from the sentence processing variant of the model. At study time, when a proposition is read, it is associated with its schema analog. At recall, it is plausible that the schema will be more active than the propositions that were studied. Moreover, those propositions in the schema which corresponded to propositions that were studied should be more active than the others in the schema and should have higher probability of retrieval. It is possible to make schema-based intrusions if there are very active schema propositions which have no equivalent in the set of studied propositions. Also, if several stories based on the same script have been studied, then the schema propositions corresponding to the union of propositions in the two texts will be more active than the rest of the schema propositions, and therefore intrusions from that set are more likely. Also, because there are more schema propositions with higher activation that were not part of the story, there will be a greater number of intrusions. Predictions obtained by such an ACT-R model implementing the recall part of this theory are shown in the last two columns of Table 12. However, since our final goal is to build a unique model that performs both comprehension and recall, the details of building the structure corresponding to the schema must be filled in.

Indeed, in order to obtain one unified, plausible model, that would feed the products of comprehension into the "recall" module, we need to reconsider our representation of background knowledge. The sentence processing model postulated the background knowledge as a repository of independent propositions; for text memory we need to deal with related propositions forming a schema. Moreover, the schema representation should be simple enough to allow activation spreading within the schema; however, they must also be sufficiently general to allow, for instance, schemas within schemas (needed for modeling Owens et al., 1979, for instance). Also, in the current text representation used by our model, there is no notion of binding: such a model would make perfect sense of a text like the following: *At the picnic Joe took some cold beer out of the car trunk. The beer was warm*, since both propositions could be found in a picnic schema. The binding problem is closely related with the problem of solving anaphoric references.

Additional problems refer to the range of the phenomena to be modeled. For instance, some of the intrusions reported by Bower et al. (1979) are name changes; i.e. subjects attribute to one character actions that were performed by another character in a related story. Also, the abstract intrusions described in Bower et al. (1979) seem hard to produce by such a simple-minded model.

A possible solution to these problems (abstract intrusions, recursive representations for schemas) may come from a more hierarchical representation of schemas, in the spirit of Schank's modified script concept (Schank, 1982, 1999). Namely, one could assume that schemas are in fact a sequence of pointers to episodes that typically occur in the situation that the schema describes. For instance, the dentist script may actually be a sequence of pointers to a getting-there prototype, to a waiting-room prototype, to a being-prepared-for examination prototype, etc. Such prototypes that describe episodes that can occur within a script may be shared among scripts. Thus, a visit to a doctor may have some episodes in common with a visit to a dentist. One reason for an abstract inference like incorrectly remembering *The nurse checked John's blood pressure and weight* when in a dentist story *The dental hygienist checked and x-rayed Bill's teeth* has been studied (as reported in Bower et al. (1979)) may be that the episode being-prepared-for-examination got more active due to having been accessed at study in the dentist story. Of course, one should still explain how the subjects report a schema specific realization of the episode (i.e. how they came to report checking blood pressure and weight). Also, such a hierarchical schema representation may help in modeling the results in Owens et al. (1979), where several episodes sharing the same character were studied by the subjects and the motivations attributed to that character influenced the pattern of recall of each episode.

While current models of text processing focus on the situation model or mental model for the text and discuss less how background knowledge comes into play (for instance, is a sentence mapped onto both background knowledge and text knowledge?), the model we proposed focuses mostly on background knowledge and ignores text coherence relations. It is a worthwhile endeavor to find an account that connects background knowledge with the situation model. Sanford and Garrod (1998)'s implicit focus/explicit focus theory may provide a good starting point.

We plan to look into these issues in more detail. One extra motivation comes from the metaphor literature, as pointed out in the metaphor understanding section. There, we have suggested that some metaphor results referring to the comprehension of anaphoric metaphors (Ortony et al., 1978; Gibbs, 1990) seem to be best discussed in a discourse processing framework. Namely, if a different time course

	No topic	Topic after	Topic before
Experiment 2	.16	.15	.32
Experiment 4	.23	.23	.4

Table 14: Recall rates in experiments 2 and 4 from Bransford & Johnson (1972).

is to be blamed for the inconsistencies between Gibbs (1990) and Ortony et al. (1978)'s results, then a theory of solving anaphoras and a description of how and when the current sentence is integrated in the text context may be needed. Moreover, since metaphors are understood based on background knowledge, anaphoric metaphor comprehension may be a good starting point for looking into the issue of how text representation and background knowledge interact in building a coherent product of comprehension.

5.2 Elaborative Encoding

In the previous section, we have reviewed a couple of studies that demonstrated how scripts can affect the memory for a specific story. While the results of these studies have been mixed, and scripts have been shown to improve but also alter the memory for the propositions actually studied, a similar line of research has indicated that presenting people with texts embellished with elaborations can actually improve the memory for those texts.

In one such study, Bransford and Johnson (1972) ran three experiments, in which they showed subjects a passage either preceded or followed by a topic (e.g. *washing clothes*) or with no topic information associated[12]. Bransford and Johnson asked subjects to rate the passage for comprehension and then to recall sentences from it. Both the comprehensibility ratings and the recall rates were significantly lower for the no-topic or topic-after conditions than for the topic-before condition[13]. The recall rates for experiments 2 and 4[14] from Bransford and Johnson (1972) are given in Table 14.

The model could account for the data from Bransford and Johnson (1972) by assuming that, in the no-topic or in topic-after condition, no unique context is attached to the sentences in the passage, although they might be interpreted individually (i.e. separate contexts may be found for each of them). However, for the topic-before condition, the model relates all the sentences to the same context and then uses the context as a cue for recalling the sentences. Indeed, if the context can be retrieved, then the propositions that have been studied have context correspondents with higher activation than non-studied context propositions, and thus they will be more readily retrieved. Thus, recalling one fact (the context or a sentence that will be related to that context) enables recalling of others. But in the no-topic and topic-after conditions, because no unique context is present, it is hard for the model to recall all the context involved, and even if it succeeds in recalling one context or proposition, that will not help with the retrieval of others.

Bradshaw and Anderson (1982) ran a series of experiments in which they examined memory for either isolated facts or for several facts about the same historical figure. If more than one fact was presented, they were all about the same figure and could be either related[15] or unrelated. Table 16 shows some examples. After an extensive study phase, in a second session happening one week after the first, Bradshaw and Anderson asked subjects to recall the facts using the names of the historical characters as cues. The recall was greater in the related condition than in the single condition. In the first two experiments, the level of practice was too high to obtain a significant difference between the related and single fact condition. However, in the third experiment this difference attained significance,

[12] The passage was written such that the topic could not be inferred from its contents.

[13] For experiment 4, there was no difference in the topic-before and topic-after conditions for comprehensibility ratings.

[14] Experiment 3 in Bransford and Johnson (1972) had no control condition, so we will not attempt to model it. Experiments 2 and 4 differed only in the passage they used and in the presentation form. Experiment 2 used a passage about washing clothes; experiment 4 used a passage about making and flying a kite and the component sentences were presented as sentences rather than in paragraph form.

[15] Bradshaw and Anderson (1982) manipulated also the type of relation, but this manipulation will not concern us here.

Related condition:
Newton became emotionally unstable and insecure as a child
This fact was caused by:
Newton's father died when he was born
Newton's mother remarried and left him with his grandfather

Unrelated condition:
Newton became emotionally unstable and insecure as a child
This fact is unrelated to:
Newton was appointed Warden of the London mint
Newton went to Trinity College in Cambridge

Single condition:
Newton became emotionally unstable and insecure as a child

Table 15: Materials used in Bradshaw & Anderson (1982). They use two related conditions (caused-by and resulted-in) which are aggregated in this discussion.

	Related	Unrelated	Single
Experiment 1	.95	.80	.93
Experiment 2	.73	.45	.61
Experiment 3	.57	.32	.38

Table 16: Recall rates from Bradshaw & Anderson (1982). The rates reported for the related condition are the means for the caused-by and resulted-in conditions reported in the original study.

with the recall in the related condition being better. The data from Bradshaw and Anderson (1982) are reported in Table 16.

As in the case of the Bransford and Johnson (1972) study, the model should be able to capture the data in Bradshaw and Anderson (1982) based on finding a long-memory situation which can serve as a context for the facts to be learned and which can be used as a basis for recall. Caution should be taken with respect to what this context can be — presumably, if the subjects had not known the facts used in these experiments (some of which were real), they should have used different background knowledge propositions than those involved in their knowledge about the historical figures. Also, one challenge in modeling these experiments is that, unlike in the Bransford and Johnson (1972) study where subjects saw one single passage and therefore could easily recall its topic in the topic-before condition, in these experiments there is possible confusion among contexts, since more than one topics were presented. The character names should play a role in cueing the recall of the correct context.

6 Syntactic Ambiguity as a Computational Constraint on the Theory

While describing the data sets on which this theory is to be tested, we also mentioned some computational constraints that it needs to satisfy. Those constraints referred to realistic reaction time predictions and to integration of new information with background knowledge. We will now discuss more extensively one such constraint — accommodating syntactic ambiguities— that has been alluded to and needs more extensive discussion.

Perhaps the most famous garden path sentence is *The horse raced past the barn fell.* What is peculiar about this sentence is that when subjects read *raced* first, they tend to believe it is the past-tense verb of the main clause; however, after reading *fell* they realize that *raced* should be parsed as an adjective rather than as a verb. The garden path refers to showing a parsing preference to one construction versus another. There are essentially two views (see, for example, Foss, 1988 for a review) of the garden-path phenomena: one is the modular view that the preferred syntactical structure respects some semantics-independent principles (like minimal attachment and late closure), the other is the interactive view that postulates that the context, as well as the meaning of the words, their frequency in various thematic roles, the argument structure of verbs etc. dictate the preferred parsing. In terms of computational theories addressing garden path phenomena (and other syntactic effects), Lewis (1993, 1999), Clark and Gibson (1988) and Gibson (1991) offer complex examples of incremental models producing concrete predictions of people's linguistic preferences. Lewis (1993) presents a variant of SOAR that tackles syntactic parsing at high level of detail, focusing less on semantic aspects of language comprehension.

A more subtle case of garden path sentences, in the sense that people may not recognize them as such at a "conscious" level, are sentences like *The cop arrested by the detective was guilty of taking bribes.* In such sentences, as McRae, Spivey-Knowlton, and Tanenhaus (1998) have shown, people experience a comprehension cost due to expecting *the cop* to be an agent of the verb *arrested.* Such an expectation is based on background knowledge — namely, *cop* is more often encountered as the agent rather than the patient of *arrest.* Therefore, while background knowledge may help with understanding *The crook arrested by the detective was guilty of taking bribes*, it hinders the comprehension of *The cop arrested by the detective was guilty of taking bribes.*

In its present form, our model relies on syntactic information (namely thematic roles) to identify the correct context; thus it performs no syntactic parsing. In doing so, it assumes that the syntactic processor is always correct. As the existence of garden path sentences shows, this assumption is wrong. However, we do not intend to include in this model a complete theory of parsing — we will confine to a model able to deal with the errors of the syntactic processor. Such a task seems to be feasible if one relaxes the condition that the thematic role be a fixed index in context identification. As the study by McRae et al. (1998) shows, the thematic role plays a role in context identification and people have expectancies about the thematic roles of noun phrases, so it would be wrong not to use thematic roles for finding a candidate context. However, the model needs to allow for variable thematic roles for the same word, that is, it needs to allow for *the cop* in *The cop arrested by the detective was guilty* to be first an agent, and then, after the correct agent (*the detective*) was read, to be reinterpreted as a patient. Such a behavior can be achieved by treating the thematic role as another feature of the meaning, which might be omitted from check by the model later on, when the initial parsing was discovered incorrect. To exemplify, when the word *the detective* in *The cop arrested by the detective was guilty* is processed, the current candidate context, which considers *arrested* as a main verb and *cop* as an agent, is abandoned and replaced with a context in which *detective* is the agent. When going back to check the features of the words that were previously processed (i.e. the words before *detective*), it is possible that the model omits to match the new context against the *agent* feature of *cop.* If the new context, in which *detective* is the agent, is accepted, *cop* will automatically be assigned the patient role. The model needs to be modified to reinforce the omitting of wrong features.

7 Summary

The thesis will describe a language comprehension theory, which should satisfy two types of constraints: computational and empirical. We will review here each of these constraints.

In the list of computational constraints, we have included three items:

Realistic reaction times: the processes subsumed by this theory should be accommodated in a time interval on the order of a few seconds.

Integration with background knowledge: the theory should account for how background knowledge is accessed during sentence and discourse comprehension. In doing so, it should also address the various intrusion types that are made in the text memory experiments (mixing-up intrusions and abstract level intrusions, especially).

Syntactic ambiguity: the model should allow for processing of garden path sentences, namely of sentences for which the initial parsing proves to be incorrect. However, we do not intend to build a complete syntactic parser: we simply plan to allow for the semantic analysis to recover from incorrect thematic role assignment.

While these constraints should provide the general guidelines for designing the theory, we intend to produce a unique ACT-R model implementing this theory[16] , to account for the following empirical data:

Metaphor understanding :

- effects of position on metaphor understanding (Gerrig & Healy, 1983);
- effects of metaphoric truth on the judgement and recall of sentences of the type *Some As are Bs* (Glucksberg et al., 1982);
- interferences of literal and metaphoric truth on sentence judgement (Keysar, 1989);
- effects of context length on metaphor understanding (Ortony et al., 1978);
- comprehension differences between different types of metaphors, suggested by comparisons between data from Gibbs (1990) and Ortony et al. (1978), and also by work in progress by John Anderson and me.

Moses illusion :

- illusion rates for good and bad distortions in the literal and gist tasks (Ayers et al., 1996);
- latencies in the literal and gist task (Reder & Kusbit, 1991);
- processing of semantic anomalies and contradictions (Barton & Sanford, 1993).

Lexical ambiguity :

- gaze durations on balanced or unbalanced ambiguous words (Duffy et al., 1988);
- mean reading time per character in the disambiguation region (Duffy et al., 1988).

Memory for text (scripts and elaborations) :

- recall and recognition of sentences from multiple episodes related or not by a common setting (Owens et al., 1979);
- interferences from related stories in recall and recognition of text (Bower et al., 1979);
- text recall in the presence or absence of a topic (Bransford & Johnson, 1972);
- recall of single, related and unrelated facts (Bradshaw & Anderson, 1982).

Even if one achieved the goal of building a model that satisfied all these constraints, this would not guarantee that it was the right model. A useful type of sanity check is collecting empirical data to validate the "side effects" of this model, that is, those predictions that are not covered by the set of empirical phenomena listed here. For instance, in the Moses illusion section we have described how this model predicts a position effect on semantic illusions. More data should be collected to support this prediction.

[16] "Unique" up to a number of internal ACT-R parameters — level of noise, base level activation values etc., which are conceivably experiment-specific.

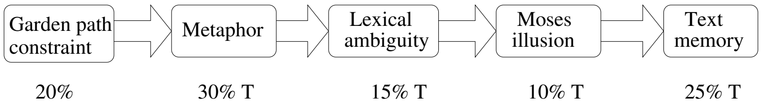

Figure 9: Work plan for the thesis and percentage of time estimated for each major task out of the total time.

Another type of validation for this model would be to test it on other sets of data (for the same phenomena) than the ones it has been built for. For instance, we can predict the Moses illusion for the question *How many animals did Moses take on the ark?* based on the feature overlap between *Moses* and the agent of the ark story. It would be useful to test the model on other items for which feature overlap has separately been established and to see whether for those items, the predictions match the real data. For such a test to be informative, it is not enough to simply encode other items in the system a posteriori using the same assumptions that were made for the *Moses* and the ark story representations — obviously, for similar encoding, the model would behave similarly. we intend to apply such a validation principle for at least some of the data sets that were described in this proposal.

8 Meta-analysis: Work Plan

In the previous sections we have proposed a model and a list of tasks to which we intend to apply the model. There are no strict dependencies in the sequence in which these tasks should be tackled, Modeling tasks which depend on data collection (like modeling possibly-different comprehension loads due to the "amount of metaphoricity" contained in a sentence or confirming the semantic illusion model by collecting similarity ratings) should, of course, be preceded by the experiments that supply that data. Other than that, we see the process of working on this thesis as trial-and-error: each time the model is extended to a new set of data, the new variant should be tested again on the old sets that it used to fit. We believe a reasonable order is the one given in Figure 9. We will start with extending the model to satisfy the syntactic-ambiguity computational constraint. Further on, we will attempt to account for the metaphor experiment that we ran last fall, hoping that it would provide a good starting point for modeling not only other metaphor-related phenomena on our list, but also the text memory data. Next, we plan to check in more detail how this model behaves on the lexical ambiguity data. Before dealing with discourse and memory for text, we will need to revise and add the missing pieces to give a complete account of the semantic illusion literature. And finally, we hope that the sentence comprehension model will be easily extended to the text memory data on our list. At this point, the most difficult part in modeling the text memory literature seems to be the building a reasonable representation for discourse.

Figure 9 also presents estimates of the time-span of the tasks. At this point, fitting the metaphor data (especially the data in Ortony et al., 1978 and the results of our metaphor experiment) looks like the most-time consuming task, especially because we have explored this direction very little. Another difficulty is making the model satisfy the garden path computational constraint: we do not estimate a large volume of modeling work to be involved in this task, but it seems more complicated because it does not arise as naturally as the other tasks from the current version of the model. With respect to lexical ambiguity, the model has already strong theoretical predictions; however, more work needs to be done in obtaining concrete predictions and comparing them with real data. Finally, the model seems to fit pretty well the existent Moses illusion data; however, we plan to gather more similarity data to add an external source of validation for our model. Also, for text memory effects, while the basic recall phenomena seem to be captured by the model, more effort is needed to replicate the more detailed aspects of data (e.g. the exact pattern of intrusions).

The majority of this work will involve modeling: both actual ACT-R programming and parameter fitting (with the aim of obtaining the best possible predictions). We expect the rest of it to be actual data collection and statistical analysis. The final purpose of this thesis (and the "termination" criterion) is to present a unique model that is able to predict data from the domains listed in the previous section.

Should such a purpose prove unattainable, the thesis will need to justify why.

We have presented some incipient ideas concerning sentence processing, discourse processing and their relationship with background knowledge. We believe that the topic of modeling sentence and discourse processing has the advantage of being open-ended, in the sense that one can calibrate the task in both depth and breadth according to time and capabilities: since, by its nature, cognitive modeling entails a number of assumptions and simplifications, it is always possible to make the modeling task more complicated or simpler by varying the number of assumptions one is willing to make. However, the additional constraint of building a unique model (up to parameters reflecting experiment-specific levels of noise or practice etc.) to fit data from multiple domains limits considerably the freedom of the modeler, while giving more credibility to the model.

References

Anderson, J., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.

Ayers, M., Reder, L., & Anderson, J. (1996). *Accepting false information now and believing it later: Partial matching and false information in the Moses illusion.* (Unpublished manuscript)

Bartlett, F. (1932). *Remembering: A study in experimental and social psychology.* New York & London: Cambridge University Press.

Barton, S., & Sanford, A. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory and Cognition*, *21*, 477-487.

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

Bradshaw, G., & Anderson, J. (1982). Elaborative encoding as an explanation of levels of processing. *Journal of Verbal Learning and Verbal Behavior*, *21*, 165-174.

Bransford, J., & Johnson, M. (1972). Contextual prerequisites for understanding: some inverstigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717-726.

Clark, R., & Gibson, E. (1988). A parallel model for adult sentence processing. In *Proceedings of the 10th annual conference of the Cognitive Science Society* (p. 270-277). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Duffy, S., Morris, R., & Rayner, K. (1988). Lexical ambiguities and fixation times in reading. *Journal of Memory and Language*, *27*, 429-446.

Erickson, T., & Mattson, M. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*, 540-552.

Foss, D. (1988). Experimental psycholinguistics. *Annual review of Psychology*, *39*, 301-348.

Gentner, D., & France, I. (1988). The verb mutability effect: Studies of the combinatorial semantics of nouns and verbs. In S. Small, G. Cottrell, & M. Tanenhaus (Eds.), *Lexical ambiguity resolutions: Perspectives from psycholinguistics, neuropsychology and artificial intelligence* (p. 343-382). Morgan Kauffman.

Gerrig, R., & Healy, A. (1983). Dual processes in metaphor understanding: Comprehension and appreciation. *Journal of Experimental Psychology: Memory and Cognition*, *9*, 667–675.

Gibbs, R. (1990). Comprehending figurative referential descriptions. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*, 56–66.

Gibson, E. (1991). *A computational theory of human linguistic processing: memory limitations and processing breakdown.* Unpublished doctoral dissertation, Carnegie Mellon University. (Available as technical report CMU-CMT-91-125.)

Glucksberg, S., Glidea, P., & Bookin, H. (1982). On understanding literal speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior*, *21*, 85–98.

Jaarsveld, H., Dijkstra, T., & Hermans, D. (1997). The detection of semantic illusions: Task specific effects for similarity and position of distorted terms. *Psychological research*, *59*, 219-230.

Kamas, E., & Reder, L. (1995). The role of familiarity in cognitive processing. In E. O'Brien & R. Lorch (Eds.), *Sources of coherence in reading* (p. 177-202). Hillsdale,NJ: Erlbaum.

Kamas, E., Reder, L., & Ayers, M. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory and cognition*, *24*, 687-699.

Kellas, G., Paul, S., Martin, M., & Simpson, G. (1991). Contextual feature activation and meaning access. In G. Simpson (Ed.), *Understanding word and sentence* (p. 47-71). Amsterdam: North Holland.

Keysar, B. (1989). On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language*, *28*, 375-385.

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

Lewis, R. (1993). *An architecturally-based theory of human sentence comprehension.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA. (Available as technical report CMU-CS-93-226.)

Lewis, R. (1998). Reanalysis and limited repair parsing: Leaping off the gardenpath. In J. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (p. 247-285). Netherlands: Kluwer Academic Publishers.

Lewis, R. (1999). Specifying architectures for language processing: Process, control and memory in parsing and interpretation. In M. Crocker, M. Pickering, & C. Clifton (Eds.), *Architectures and mechanisms for language processing.* Cambridge University Press.

Lucas, M. (1999). Context effects in lexical access: a meta-analysis. *Memory and Cognition*, *27*, 385-398.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676-703.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*, 440-466.

McRae, K., Spivey-Knowlton, M., & Tanenhaus, M. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*, 283-312.

Onifer, W., & Swinney, D. (1981). Accessing lexical ambiguity during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition*, *9*, 225–236.

Oostendorp, H. van, & Kok, I. (1990). Failing to notice errors in sentences. *Language and cognitive processes*, *5*, 105-113.

Oostendorp, H. van, & Mul, S. de. (1990). Moses beats Adam: A semantic relatedness effect on a semanticillusion. *Acta Psychologica*, *74*, 35-46.

Ortony, A., Schallert, D., Reynolds, R., & Antos, S. (1978). Interpreting metaphors and idioms: Some effects on comprehension. *Journal of Verbal Learning and Verbal Behavior*, *17*, 465–477.

Owens, J., Bower, G., & Black, J. (1979). The "soap opera" effect in story recall. *Memory and Cognition*, *7*, 185-191.

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*, 191-201.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 779-790.

Reder, L. (1980). The role of elaboration in the comprehension and retention of prose: A critical review. *Review of Educational Research*, *50*, 5–53.

Reder, L., & Kusbit, G. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, *30*, 385-406.

Sanford, A., & Garrod, S. (1998). The role of scenario mapping in text comprehension. *Discourse processes*, *26*, 159-190.

Schank, R. (1982). *Dynamic memory: A theory of reminding and learning in computers and people.* Cambridge: Cambridge University Press.

Schank, R. (1999). *Dynamic memory revisited.* Cambridge: Cambridge Universit Press.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Schustack, M., & Anderson, J. (1979). Effects of analogy to prior knowledge on memory for new information. *Journal of Verbal Learning and Verbal Behavior*, *18*, 565-583.

Searle, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and thought.* Cambridge University Press.

Simpson, G. (1994). Context and the processing of ambiguous words. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (p. 359-374). San Diego, CA: Academic Press.

Swinney, D. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645-659.

Tabossi, P. (1988). Accessing lexical ambiguities in different types of sentential context. *Journal of Memory and Language*, *17*, 324-340.

Webster's 9th new collegiate dictionary. (1991). Springfield, MA: Merriam-Webster.

Zwaan, R., & Radvansky, G. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162-185.