

Effects of Discourse Context on Inference Computation during
Text Comprehension

by

Yasunori Morishima

and

Walter Kintsch

Department of Psychology
and
Institute of Cognitive Science
University of Colorado, Boulder

RUNNING HEAD: Discourse Context Effects on Inferences

Send correspondence to:
Yasunori Morishima
10251-A Rancho Place
Cupertino, CA 95014
(408) 366-0545 (home)
(650) 578-8067 (work)
(650) 572-7108 (fax)
ymorishima@dyned.com

Abstract

Comprehenders construct a situation model of a text that they are processing. The situation model is a microworld of agents, objects, actions, events, spatial composition, and states associated with what the text explicitly asserts. The current study asks how a situation model influences inference computation during comprehension. The primary question is how strongly a comprehender infers a high-probability instrument (e.g., a pump for inflating a tire) during comprehension given a brief text. The experiments showed that instrument inferences are highly affected by the discourse context; such inferences are not computed unless the context provides information that strongly supports the inferences (e.g., "John grabbed the pump in the garage."). When the context is compatible with such inferences but weak (e.g., "John found the pump in the garage."), comprehenders do not routinely draw the inference. However, comprehenders do infer instruments when they are motivated to elaborate during comprehension. A simulation model was developed based on Kintsch's (1988) Construction-Integration theory. The model successfully displayed behaviors that are qualitatively in agreement with the experimental results and provides an explicit account for the on-line inference computation process.

1. INTRODUCTION

Consider the following sentence pairs:

- 1.1 Marvin turned on the light in the shed. He inflated the tire.
- 1.2 Marvin found the broken pump in the shed. He inflated the tire.
- 1.3 Marvin grabbed the pump in the shed. He inflated the tire.

Would comprehenders infer that Marvin used the pump to inflate the tire on-line, that is, during comprehension? The first sentences in these pairs set up the situation for the actions described by the second sentences. Let us call the first sentences context sentences and the second ones action sentences. While the action sentence in each pair is the same, each context sentence implies a different situation about availability of the instrument "pump". The context sentence for Text 1.1 does not mention "pump" at all. Text 1.2 implies that the pump was not usable since it was broken, thus it does not lead to the expected inference. In other words, it would not make sense to assume that Marvin used the broken pump to inflate the tire. Text 1.3, on the contrary, suggests that he took the pump and intuitively supports the inference. How is the inference process, or specifically the activation of the high-probability instrument, affected by the context sentences? This is the primary question that we address in this article.

Instrument inference has been studied since the early days of inference research. Although early studies (e.g., Johnson, Bransford, & Solomon, 1973; Paris & Lindauer, 1976) claimed that an implicit instrument is encoded in reading a sentence such as *The lawyer cooked dinner* (the instrument is a *stove*), there is considerable evidence today that an instrument is inferred only when certain conditions are met. An explanation for it is that such an inference, called an elaborative inference, only provides additional information rather than establishing

local coherence of a text. Inferences that performs the latter function are called bridging inferences. It is generally agreed in the literature that bridging inferences normally accompany comprehension (Corbett, 1984; Corbett & Chang, 1983; Dell, McKoon, & Ratcliff, 1983; Gernsbacher, 1989; Graesser, & Kreuz, 1993; Haviland & Clark, 1974; Kintsch & Vipond, 1978; Lesgold, Roth, & Curtis, 1979; McKoon & Ratcliff, 1980, 1986) whereas elaborative inferences are not routinely computed on-line unless the context of a text is highly constraining or are encoded into the representation only "minimally" (Anderson & Ortony, 1975; Fincher-Kiefer, 1995; Keefe & McDaniel, 1993; Mauener, Tanenhaus, & Carlson, 1995; McKoon & Ratcliff, 1986, 1989; Murray, Klin, & Myers, 1993; O'Brien, Shank, Myers, & Rayner, 1988; Potts, Keenan, & Golding, 1988; Singer & Ferreira, 1983; Whitney, Ritchie, & Crane, 1992).

Several studies show that unless an instrument has been introduced in the text as in Text 1.1, the instrument inference would not be drawn on-line, and the instrument would not be activated (Corbett & Doshier, 1978; Doshier & Corbett, 1982; Singer, 1979, 1980). McKoon and Ratcliff (1981) showed that the instrument is activated at the time of the reading a sentence mentioning the action if the action is highly associated with the instrument. The researchers used texts as the following:

- 1.4. Bobby got a saw, hammer, screwdriver and square from his toolbox.
- 1.5a. Then Bobby pounded the boards together with nails.
- 1.5b. Then Bobby stuck the boards together with glue.

The passages used in the experiments included a combination of Text 1.4 and either Text 1.5a or 1.5b with a few intervening sentences. They found that the reaction time for recognition of the test word "hammer" was shorter when Text 1.5a was presented than when Text 1.5b was presented. On the other hand, when the test word was "mallet" (in this case, "mallet" is included in Text 1.4), there was no

significant difference in reaction time between the two versions. Lucas, Tanenhaus, and Carlson (1990), using the lexical decision task, obtained similar results. Namely, instrument inferences are drawn if previous discourse has explicitly mentioned the instrument, but not if sentences with actions implying the use of an instrument are presented out of context. Furthermore, van Meter and Pressley (1994) also found supporting evidence with 10- to 14-year-old children.

To summarize these studies, instrument inference takes place during comprehension under two conditions. The first condition is semantic association between an instrument and an action; namely, an action should have a strong association with a particular instrument (e.g., pounding -- hammer; driving -- car). The other condition has to do with the structure of the text; that is, an instrument should be explicitly mentioned in the text prior to the mention of an action.

However, in these studies, discourse context effects were not systematically investigated. Either the degree of relatedness between the context sentence and the action sentence was not controlled, or context effects were examined only in long-term memory representation rather than on-line processes (see McKoon & Ratcliff, 1981). Thus, the question arises here; are those conditions stated above sufficient for activating the instrument, or does discourse context influence the inference computation?

We argue that the global structure or situation model of text influences on-line computation of such inferences. In the present study, we focus on instrument inferences like the above examples. A situation model is a microworld of agents, objects, actions, events, spatial composition, and states associated with what the text explicitly asserts (Johnson-Laird, 1983; Kintsch, 1988; van Dijk & Kintsch, 1983). The view that a situation model plays an important role during comprehension has been put forward by several theories (Collins, Brown, & Larkins, 1980; Johnson-

probability instruments. Each experiment measured reaction times for instrument priming and reading times for processing critical sentences. Experiment 1 was conducted to establish that the experimental setting employed by the present study was sensitive enough to detect inference, independent of associative priming. In Experiment 2, we manipulated the degree of relatedness between the context sentence and the action sentence to examine its effects on inference computation. In Experiment 3, effects of depth of processing on inference computation was tested in effort to demonstrate evidence for a chain of inferences proposed above. In the second part, we present a computational model based on Kintsch's (1988) Construction-Integration Model. By developing the computational model, we were able to specify more explicitly the memory structure, control mechanisms, representation structure, and processes involved in inference computation.

2. EXPERIMENTS

In the present study, three experiments were conducted. Prior to the experiments, a series of pilot studies were conducted to validate empirically the experimenter's intuitions with regard to the qualities of the texts for the subsequent experiments.

Pilot Studies

Forty-nine sets of experimental texts like the example below were created with action-instrument associations based on Corbett and Doshier (1978), and Lucas et al. (1990). For each of the action sentences, four levels of context sentences were prepared and labeled as follows: Contradictory, Weakly Related, Moderately Related, and Strongly Related. The contradictory context was intended to provide the context that did not lead to the use of the instrument mentioned by the sentence. The three related context sentences were designed to provide the context in which the use of the instrument was plausible, but varied in terms of the degree of supporting the

implication. As the labels suggest, the strongly related context intuitively supported the use of the instrument most strongly. The other related contexts, on the other hand, do not readily provide such support, and would require the longer chain of events in order to establish the availability of the instrument for use.

Context Sentences:

- (Contradictory) Marvin found the broken pump in the shed.
- (Weakly Related) Marvin searched for the pump in the shed.
- (Moderately Related) Marvin spotted the pump in the shed.
- (Strongly Related) Marvin grabbed the pump in the shed.

Action Sentence: He inflated the tire.

A series of rating and ranking experiments were conducted on each version of the texts (Myers, Shinjo, & Duffy, 1987). Undergraduates at the University of Colorado participated in the experiments for course credit. They were all native speakers of English. Participants rated each context sentence, using a 7-point rating scale, on how strongly it was related to the action sentence. Similarly, in the ranking experiments, they rank ordered the context sentences.

Twenty-four best texts were selected and analyzed. The mean rating scores and standard deviations for all context levels across texts are shown in Figure 1. An analysis of variance with items as a random factor was performed on the data. The analysis was based on mean rating scores for each text for each context level. There was a significant monotonic trend of the context level [$F(1, 23) = 1788.22, p < 0.001$]. Therefore, the results established that the context sentences were well separated and were ordered according to the hypothesized degree of relatedness to the action sentence.

Insert Figure 1 about here.

Experiment 1

The purpose of Experiment 1 was to establish that the present experimental paradigm was suitable for measuring priming effects due to inference, but not to word-based associative priming because, as discussed above, priming effects in inference experiments may be contaminated with the latter type of inference. The experiment employed self-paced reading and a lexical decision task. There were two dependent variables in the experiment. One is the reaction time for the lexical decision, and the other the reading time of the action sentence.

Method

Participants

Fifty-six undergraduates from the University of Colorado at Boulder participated in the experiment for credit in an introductory psychology course. All participants were native speakers of English.

Materials and Design

There were three levels of context in this experiment: Related, Contradictory, and Unrelated. The context was a within-participant independent variable. There were 18 experimental texts, which were randomly chosen from the set of the 24 experimental texts discussed above. Each text was assigned to one of the three context conditions for each participant. A sentence pair was composed of the context sentence followed by the action sentence. Shown below is an example of the text:

2.1a. Marvin grabbed the pump in the shed.

2.1b. Marvin found the broken pump in the shed.

2.2a. He inflated the tire.

2.2b. He flew to Boston.

The sentence pair for the related context consisted of Sentences 2.1a and the action sentence 2.2a. The contradictory condition contained Sentences 2.1b and 2.2a. The sentence pair for the unrelated control context was constructed by pairing Sentence 2.1a and the action sentence from another set (Sentence 2.2b). In this case, care was taken so that the gender/number agreement, and hence referential coherence, was maintained in the resulting sentence pair. The targets for the lexical decision were instrument words mentioned in the context sentence ("pump"). Note that in all conditions, the target word appeared in the context sentence to control for the repetition effect on the lexical decision task. There were 42 filler texts, 30 of which had non-word targets that provided "no" responses for the lexical decision task. For the other 12 sentence pairs, nouns other than the instrument words from the context sentences ("shed") were used as lexical decision targets. The experimental texts were randomly assigned to the conditions for each participant. The order of text presentation was randomized for each participant. The comprehension questions after the lexical decision task asked aspects of the texts other than the use of the instrument ("Did Marvin go into the shed?").

There were two levels (250 msec and 500 msec) of stimulus onset asynchrony (SOA), which refers to the length of delay time between the offset of the action sentence and the onset of the lexical decision. This was a between-participant variable.

There are three possible outcomes of this experiment. First, if the lexical decision is dominated only by the memory trace of the instrument word remaining in the short-term memory, there should not be any difference due to the context

condition. Second, if the effect is due to semantic association, both Related and Contradictory contexts should produce a priming effect because their context sentences explicitly mention the instrument, followed by the same action sentence. Finally, if the priming effect is due to the inference about the instrument, only the related context condition should produce the effect.

Apparatus

A Macintosh Quadra 605 was used for the experiment. A program was written in C to control the experiment.

Procedure

Participants were randomly assigned to either of the two SOA conditions. They were tested individually. First the participants were given the instructions on the computer screen. Then, the instructions were further explained by the experimenter. At the beginning of each trial, a string of plus signs with the same length of the context sentence appeared on the computer screen, which was soon replaced by the context sentence. The participants read the context sentence and hit the designated key called the "next" key (the down-arrow key on the Macintosh extended keyboard) to request the action sentence. Then the context sentence was replaced by the action sentence. The participants read the sentences at their own pace. When the participant hit the "next" key after reading the action sentence, a plus sign appeared in the center of the screen for either 250 msec or 500 msec, and it was replaced by the target, to which the participants were instructed to respond by pressing either the "yes" key (the right-arrow key) or the "no" key (the left-arrow key). Then, the participants responded to a yes/no question which was presented on the screen after the lexical decision task. The participants used only an index finger to make a response. They were instructed to place their index finger on the "next" key while reading and move it when they made a response. They were also

instructed to make a response as quickly and accurately as possible. Reaction time and accuracy for lexical decision were measured. Reading time of the action sentence and accuracy for the comprehension were also measured. There was a practice run with 10 trials before the test run. In the practice run, the participants were given feedback about their reaction times for the lexical decision task and responses to yes/no questions. In the test run, no feedback was given to the participant. The first 4 trials in the test run were not experimental trials. The participants were encouraged to take a short break after a trial whenever they felt tired. The entire experiment took approximately 30 minutes for each participant.

Results and Discussion

Reaction Times for Lexical Decision on Instruments.

There were 26 participants assigned to the 250 msec SOA condition and 30 participants in the 500 msec SOA condition. Errors occurred either when a participant made an incorrect lexical decision or failed to respond within 2 sec. They were eliminated from the analyses. Reaction times greater or less than 3.0 standard deviations from the participant mean were replaced by this cut-off value. Two participants were eliminated from the analyses due to high error rates. Mean reaction times for all conditions are reported in Table 1. This treatment was performed for the other experiments.

Insert Table 1 about here.

Analyses of variance with participants and items as random factors were performed on the reaction time data. All analyses were based on mean reaction times for each participant or item in each condition. They are reported here as F_p

and F_i respectively. The main effect of SOA was not significant by the participant analysis [$F_p(1, 53) = 0.21, p > 0.6$], but was significant by the item analysis [$F_i(1, 17) = 8.23, p < 0.02$]. However, the SOA was not shown to have any influence on the other effects, as no interactions with other variables were significant by either analysis [$F_s < 1.5$].

The mean reaction time for the related condition was significantly faster than those for the contradictory and unrelated conditions by the participant analysis [$F_p(1, 53) = 12.66, p < 0.001$]. This effect was marginally significant by the item analyses [$F_i(1, 17) = 3.72, 0.07 < p < 0.071$]. Planned pair-wise comparisons of the related context with the unrelated and the contradictory conditions were also performed. The reaction time for the related context was significantly faster than each of these context conditions by the participant analysis [$F_p(1, 53) = 9.85, p < 0.003$; $F_p(1, 53) = 10.16, p < 0.003$]. The item analyses showed that this effect was marginally significant with respect to the unrelated context [$F_i(1, 17) = 3.72, 0.08 < p < 0.081$], but it did not reach significance with respect to the contradictory context [$F_i(1, 17) = 2.58, p > 0.1$]. The contradictory condition did not differ significantly from the unrelated condition by either analysis [$F_p(1, 53) = 0.03, p > 0.8, F_i(1, 17) = 0.03, p > 0.8$]. Strictly speaking, the priming effect for the related context condition did not reach statistical significance in the item analyses, thus we note that we should be cautious about generalizing the effect over various texts. Yet, considering the fact that the item analysis did not have as much power as the participant analysis and that the effect was close to significance, it does not seem unreasonable to assume that the effect would be found with different texts.

The results showed that there was instrument priming after reading the action sentence if the discourse context provided the information that supported the inference, but there was no such priming if the context implied that the instrument

was not available for use. Since the context sentence contained the instrument word in both the related and contradictory conditions, if the priming effect were due to semantic association between the action and the instrument alone, both conditions should have produced the priming effect. Thus, the results suggest that the instrument priming observed in the experiment was due to the instrument inference, not an associative priming.

Reading Times of Action Sentence.

The reading time data for the action sentence obtained from the 54 participants were analyzed. Reading times greater than 5 sec were eliminated from the analyses. In addition, reading times greater than 3 standard deviations from the participant mean were replaced by this cut-off value. This treatment was performed for the other experiments. The mean reading time for each condition is shown in Table 2.

Insert Table 2 about here.

Analyses of variance with participants and items as random factors were performed on the data. All analyses were based on mean reading times for each participant or item in each condition. They are reported here as F_p and F_i respectively. The main effect of SOA was not significant by the participant analysis [$F_p(1, 53) = 0.96, p > 0.3$], but was significant by the item analysis [$F_i(1, 17) = 18.3, p < 0.001$]. However, no interactions with other variables were found to be significant by either analysis [$F_s < 0.29$].

The mean reading time for the related condition was significantly faster than those of the contradictory and unrelated conditions by both analyses [$F_p(1, 53) = 25.33, p < 0.001, F_i(1, 17) = 18.30, p < 0.001$]. Planned paired comparisons of the

related context with the unrelated and the contradictory conditions were also performed. The reading time for the related context was significantly faster than each of these context conditions by the participant analysis [$F_p(1, 53) = 12.82, p < 0.001$; $F_p(1, 53) = 22.8, p < 0.001$]. The item analyses also showed that this effect was significant with respect to both contexts [$F_i(1, 17) = 6.74, p < 0.02$; $F_i(1, 17) = 20.32, p < 0.001$]. The contradictory condition did not differ significantly from the unrelated condition by either analysis [$F_p(1, 53) = 0.24, p > 0.6$, $F_i(1, 17) = 0.11, p > 0.7$].

The reading time data showed the same pattern of results as the reaction time data. If the event described by the action sentence did not fit the context and thus was not coherent at the discourse level (e.g., pounding the board in the context of setting up the projector in the lab or finding the broken hammer), the participants took a longer time to process the action sentence. This outcome suggests that the participants tried to elaborate on the text to establish global coherence of the text. Note that the referential coherence was maintained for all the contexts by the use of the proper pronoun for the agent in the action sentence. Yet, the unrelated and contradictory conditions showed significantly longer reading times than the related condition. If coreference or argument overlap is the major source of coherence that comprehenders attempt to maintain during comprehension, there should be no difference in reading time among the three conditions.

Experiment 1 successfully established that the present method is sensitive to inference but not to word-based associative priming. The reaction time data and the reading time data provided converging evidence for this conclusion. These results give support to the hypothesis that global structure of text affects on-line processes, in particular that computation of instrument inference depends on discourse context. However, it is not clear how much elaboration comprehenders would make to draw instrument inference because in Experiment 1 the context sentences

in the related condition were not only consistent with the instrument inference but also strongly related to the action. The next experiment was conducted to address this issue.

Experiment 2

Experiment 2 tested the hypothesis concerning the construction of an inference path leading to instrument inference. The experiment tested whether an instrument inference would be computed even if the discourse context calls for an elaboration to infer an instrument.

Method

Participants.

The participants were 36 native speakers of English in the Boulder community, who were paid for their participation.

Materials and Procedure.

Experiment 2 was identical with Experiment 1 for the most part. The only difference was the context condition. In this experiment, there were four levels of context: Unrelated, Weakly Related, Moderately Related, and Strongly Related. An example of the text is shown below. The sentence pair for the unrelated context was constructed in the same manner as in Experiment 1. The three related contexts differ in the degree of relatedness to the action sentence as discussed in the pilot study section.

Context Sentences:

- (Weak) Marvin searched for the pump in the shed.
- (Moderate) Marvin spotted the pump in the shed.
- (Strong) Marvin grabbed the pump in the shed.

Action Sentence: He inflated the tire.

There were a total of 24 experimental texts, thus 6 texts for each context. There were 36 filler texts, 30 of which had non-words as targets to provide "no" responses, and the other 6 texts were positive fillers that had words from the context sentence other than instrument words as targets. The SOA of 500 msec was used in this experiment because no interactions of SOA with any other effects were found in Experiment 1.

Results and Discussion

Mean reaction times for the lexical decision task and mean reading times for the action sentences are presented in Table 3.

Insert Table 3 about here.

Reaction Times for Lexical Decision on Instruments.

Analyses of variance with participants and items as random factors were performed on the reaction time data. All analyses were based on mean reaction times for each participant or item in each condition. The mean reaction time for the strongly related condition was significantly faster than those for the other context conditions [$F_p(1, 35) = 24.47, p < 0.001, F_i(1, 23) = 22.12, p < 0.001$]. Planned pair-wise comparisons of the strong context condition were also performed. The mean reaction time for the strong context was significantly faster than each of the other context conditions by both analyses [against Unrelated, $F_p(1, 35) = 19.56, p < 0.001, F_i(1, 23) = 26.46, p < 0.001$; against Weak, $F_p(1, 35) = 13.47, p < 0.001, F_i(1, 23) = 7.33, p < 0.001$; against Moderate, $F_p(1, 35) = 12.32, p < 0.002, F_i(1, 23) = 5.73, p < 0.003$]. Though the mean reaction time for the unrelated context was approximately 30

msec slower than those for the weak and moderate contexts combined, this difference was not significant by either analysis [$F_p(1, 35) = 2.05, p > 0.16, F_i(1, 23) = 0.70, p > 0.41$]. This means that there was no significant instrument priming for either of the weaker contexts. It was also found that the mean reaction times for the three related contexts were, on average, significantly faster than the mean reaction time for the unrelated context [$F_p(1, 35) = 6.87, p < 0.02, F_i(1, 23) = 4.91, p < 0.04$]. This result replicated what had been found in the literature (e.g., Lucas, et. al., 1991; McKoon & Ratcliff, 1981). Previous studies did not control for the degree of relatedness between the context and action sentences and obtained the evidence that supported instrument priming. What this analysis suggests is that the instrument priming found in the previous studies may in fact be due to a particular type of texts, namely the texts for the strongly related condition.

The reaction time data indicates that the instrument inference was computed only when the discourse context provides information that strongly supports such an inference. Thus, the result leads to the interpretation that comprehenders do not routinely engage in an extensive elaboration to construct a long inference path leading to the instrument inference. In fact, in light of the fact that only the strongly related condition yielded the instrument priming, comprehenders integrate only a minimum amount of knowledge during comprehension.

Reading Times of Action Sentence.

Analyses of variance with participants and items as random factors were performed on the reaction time data. All analyses were based on mean reading times for each participant or item in each condition. The mean reading times for the three related contexts were, on average, significantly faster than that for the unrelated context [$F_p(1, 35) = 38.41, p < 0.001, F_i(1, 23) = 9.94, p < 0.005$]. The reading time for the strongly related context was approximately 64 msec. faster than those for

the other related contexts. This difference was marginally significant in the participant analysis [$F_p(1, 35) = 3.22, 0.081 < p < 0.082$], but did not reach significance in the item analysis [$F_i(1, 23) = 1.85, p > 0.18$]. Planned pair-wise comparisons of these related contexts with the unrelated context were also performed. Each of these conditions yielded significantly faster reaction times than the unrelated condition by both participant and item analyses [for Weak, $F_p(1, 35) = 15.19, p < 0.001, F_i(1, 23) = 6.23, p < 0.03$; for Moderate, $F_p(1, 35) = 26.33, p < 0.001, F_i(1, 23) = 7.06, p < 0.02$; for Strong, $F_p(1, 35) = 61.24, p < 0.001, F_i(1, 23) = 12.26, p < 0.002$].

As in Experiment 1, the reading time data showed the coherence effect. However, this disadvantage in processing time was found only for the unrelated and contradictory conditions. In Experiment 2, the mean reading times for the weaker contexts were significantly faster than that for the unrelated condition and were not very different from the reading time for the strong context condition. This suggests that whereas the unrelated and contradictory contexts build a situation model with which the subsequent sentence cannot be integrated, the weaker contexts, though not strong enough to generate the inference, are good enough to set up a situation model that is compatible with the subsequent action sentence. The faster reading time for these contexts also suggests that the texts for the weak and moderate conditions were processed in a very similar manner as those for the strong condition.

Why did only the strongly related context yield the instrument priming? One explanation is that the strong context does not require deep processing to generate the instrument inference because information or knowledge necessary for the inference can be retrieved by the text easily. Namely, 'grabbing the pump' also entails 'having it', and this information and the association between 'pump' and 'inflating (a tire)' in the situation model make it easier to draw the inference. On

the other hand, the weaker contexts do require more elaborations because such information is not easily retrieved by the text elements. This explanation leads to the hypothesis that if the participant does engage in deeper processing, these weaker contexts should also show the instrument inference. Experiment 3 tested this hypothesis.

Experiment 3

Experiment 2 found instrument priming only for the strongly related context. The reading time data suggested that there was not a significant difference among the three related conditions in terms of processing time of the action sentence. It is hypothesized that the comprehender devotes only a limited amount of cognitive resource for on-line inference processing unless there is a reason for deeper and elaborative processing. If, on the other hand, the comprehender engages in the deeper processing that activates and integrates more general knowledge, instrument inference should be drawn. This experiment tested this hypothesis by having the participant engage in a strategic elaborative reading. It was assumed that if the hypothesis holds and participants indeed engage in deeper processing of the text, the instrument would be inferred even for the weaker context.

Method

Participants.

The participants were 36 native speakers of English in the Boulder community, who were paid for their participation.

Materials, Design, and Procedure.

This experiment was almost identical to Experiment 2 in that it involved reading of a sentence pair followed by a lexical decision task and comprehension test. The present experiment had three levels of context: Contradictory, Moderately

Related, and Strongly Related. An example of the experimental texts is shown below:

Context Sentences:

(Contradictory) Marvin found the broken pump in the shed.

(Moderate) Marvin spotted the pump in the shed.

(Strong) Marvin grabbed the pump in the shed.

Action Sentence: He inflated the tire.

The difference between the present experiment and the other two experiments is that the comprehension questions were about the instrument words mentioned in the context sentences. For instance, the question for the above example was *Did Marvin use the pump to inflate the tire?* The correct answers to the questions were to be positive for all but the contradictory context. The participant was instructed to pay particular attention to each comprehension question and feedback from the computer during the practice session. The rationale for this manipulation is that it is expected that after several trials, the participant would be led to the interpretation that the instrument was used to perform the action as long as the context sentence implied that that was plausible, thus resulting in drawing the instrument inference while reading the text.

There were 18 experimental texts, which were randomly chosen from the set of texts used for Experiment 2. Thus, each context condition had 6 texts for each participant. There were 30 filler texts, 24 of which had non-words as targets to provide "no" responses, and the other 6 texts were positive fillers. There were an equal number of texts that provided "yes" responses and "no" responses to the comprehension question. That is, 6 of the experimental texts assigned to the contradictory condition were for "no" responses while the other 12 texts were for

"yes" responses. Thus, 18 filler texts were assigned to the contradictory condition to provide "no" responses to the comprehension question and 12 were assigned to the related context conditions to provide "yes" responses.

Results and Discussion

Table 4 shows mean reaction times for the lexical decision task, mean reading times for the action sentences, and mean response times for the comprehension questions.

Insert Table 4 about here.

Reaction Times for Lexical Decision on Instruments.

Analyses of variance with participants and items as random factors were performed on the reaction time data. All analyses were based on mean reaction times for each participant or item in each condition. The mean reaction times for the strongly related and moderately related conditions were, on average, significantly faster than the reaction time for the contradictory context condition in the participant analysis [$F_p(1, 35) = 8.09, p < 0.008$] and this difference was marginally significant in the item analysis [$F_i(1, 17) = 3.14, 0.09 < p < 0.095$]. The difference between the strongly related context and the moderately related context did not reach significance by either analysis [$F_p(1, 35) = 0.49, p > 0.48, F_i(1, 17) = 0.58, p > 0.45$].

Reading Times of Action Sentence.

Analyses of variance with participants and items as random factors were performed on the data. All analyses were based on mean reading times for each participant or item in each condition. Although the reading time for the contradictory context was slower by approximately 100 msec, this difference did not

reach statistical significance by either analysis [$F_p(1, 35) = 2.23, p > 0.1$; $F_i(1, 17) = 2.85, p > 0.1$]. The difference between the strong and moderate contexts was not significant by either analysis [$F_p(1, 35) = 0.003, p > 0.95$; $F_i(1, 17) = 0.000, p > 0.99$]. The results showed that the participants spent about the same amount of time on reading the action sentence regardless of context.

Response Times for Comprehension Question.

In Experiment 3, the response time for comprehension was also measured. Analyses of variance with participants and items as random factors were performed on the data. All analyses were based on mean reading times for each participant or item in each condition. The main effect of context was not significant in either analysis [$F_s < 1.13$]. Planned comparisons showed only that the difference between the moderately related context and the strongly related context was marginally significant in the participant analysis [$F_p(1, 35) = 3.86, 0.057 < p < 0.058$]. No other comparisons were significant [$F_s < 1.3$].

The reaction time data showed that there was the instrument priming for both the moderate and strong context conditions whereas Experiment 2 found instrument priming only for the strong context. Since the only difference between these experiments was that the present experiment employed the training session to attend to the use of the instrument, this result seems to be due to this manipulation. It also suggests that the participants engaged in more elaborative processing in Experiment 3. These interpretations lead to the conclusion that the instrument was inferred for the weaker context when the participants were motivated to elaborate the text.

The reading time and answer time data provide supporting evidence for this conclusion. The reading time data showed that the participants took approximately the same amount of time for reading the action sentence regardless of context. This

result can be taken to mean that their reading was more strategic and elaborative in the present experiment than in the previous experiments. The answer time data showed that there was not a major difference in the time that the participants needed to answer the comprehension questions across the context conditions. This means that all they needed to do to make a response to the question was to access the elaborated representation of the text, rather than to compute an inference at the time of the test. Experiments 1 and 2 convincingly showed that the instrument was activated after the reading of the action sentence for the strong context condition. Thus, the answer time for the strong context condition is taken to show the time required to access the memory, and the same argument should be applied to the reading time data for the moderate context.

Comparison of Experiment 3 with the other experiments

The results of Experiment 3 are interpreted to show that the participants adopted a more elaborative processing strategy due to the training and that as a result the instrument was inferred for both the moderate and strong contexts. To further examine this interpretation, the reaction time and reading time data of Experiment 3 were compared with those of the other two experiments.

Comparison of Reaction Times for Lexical Decision with Experiment 2

The reaction time data for the moderate and strong conditions of Experiments 2 and 3 were analyzed (see Figure 2). Analyses of variance with participants and items as random factors were performed on the data. All analyses were based on mean reaction times for each subject or item in each condition. Note that Experiment 2 used 24 texts whereas Experiment 3 used only 18. The item analysis used the data only for those 18 texts that were used for both experiments. There were two independent variables: Experiment was the between-participant variable and Context was the within-participant variable. Both Experiment and Context were

within-item variables in the item analysis. There was the main effect of Experiment, which was marginally significant by the participant analysis and was significant by the item analysis [$F_p(1, 71) = 2.88, .09 < p < .095; F_i(1, 17) = 16.46, p < .001$]. There was also the significant main effect of Context by both analyses [$F_p(1, 71) = 8.78, p < .005; F_i(1, 17) = 3.50, p < .008$]. Of main interest in this analysis is the interaction between Experiment and Context. It was significant by the participant analysis [$F_p(1, 71) = 4.432, p < .04$], but the item analysis did not show a significant effect [$F_i(1, 17) = .77, p > .3$]. Note that the item analysis had much less power compared to the participant analysis and thus it is not surprising that a significant interaction was not found. Therefore, although we should be cautious about generalizing this interpretation over texts, the comparison analysis provided supporting evidence for the conclusion that in Experiment 3, due to the training, instrument inference was computed not only for the strong context but also the moderate condition.

Insert Figure 2 about here.

Comparison of Reading Times for Action Sentence with Experiment 1

The reading time data of Experiment 3 were compared with those of Experiment 1. In this comparison, the data for the contradictory and strong conditions were used (see Figure 3). The lack of difference in reading times for the action sentence in the last experiment is assumed that it showed a change in the comprehension process. In particular, in light of the reaction time data, the result suggests that the participants engaged in more elaborative processing in Experiment 3 due to the training.

Insert Figure 3 about here.

Analyses of variance with participants and items as random factors were performed on the data. All analyses were based on mean reaction times for each subject or item in each condition. The experiment was the between-participant variable and the context was the within-participant variable for the participant analysis. Both experiment and context were within-item variable for the item analysis. The main effect of Experiment was marginally significant by the participant analysis [$F_p(1, 89) = 3.07, .08 < p < .09$], but was significant by the item analysis [$F_i(1, 17) = 6.95, p < .02$]. The main effect of Context was significant by both analyses [$F_p(1, 89) = 9.33, p < .005; F_i(1, 17) = 13.68, p < .002$]. That is, over all, the reading times were faster for the strong context than for the contradictory context. Like the comparison of reaction time data, of main interest is the interaction between Experiment and Context. The participant analysis showed that the interaction was marginally significant [$F_p(1, 89) = 2.972, .08 < p < .09$] and was significant by the item analysis [$F_i(1, 17) = 4.56, p < .05$].

The analysis confirmed that there was a change in the participants' comprehension behavior in Experiment 3: They spent the same amount time in processing the action sentence for both the contradictory and strong contexts. This result alone can be interpreted either as showing that in Experiment 3 the comprehenders adopted "shallow" processing for all the conditions or as showing that the participants applied "deeper" processing to all the conditions. The latter interpretation seems more plausible for two reasons. First, the training is thought to have encouraged more anticipatory, strategic comprehension of the texts because all the comprehension questions were about instrument use and the participants

were trained to think whether a given context was plausible for instrument use. Second, the reaction time data showed instrument priming for the moderate context, which suggests that the manner of comprehension was more elaborative in Experiment 3 than in Experiment 2 because the same context did not yield the priming effect in the second experiment.

Therefore, the comparison analyses of the reaction time and reading time data provided further support for the conclusion that the participants adopted elaborative processing in Experiment 3 because of the training. In other words, comprehenders do not routinely engage in an extensive elaboration to generate inferences unless they are motivated to do, but if they do engage in deeper processing, weaker contexts will also yield the inference.

3. SIMULATION MODEL

The Construction-Integration Model

In this section we present a computational model that accounts for the experimental results. The model has been designed based on the Construction-Integration (CI) theory proposed by Kintsch (1988, 1992; Kintsch & Welsch, 1991). The Construction-Integration theory provides a general architecture of comprehension processes. In the theory, comprehension arises from an interaction and fusion between the to-be-comprehended object (e.g., a text), and the world knowledge and episodic memory that the comprehender brings to the situation. The CI theory consists of two major components: a bottom-up construction and a top-down integration. During the construction phase, a set of symbolic rules of syntax and semantics, as in a production system (e.g., Newell, 1973), construct a network of propositional nodes. Propositional representations of text are elaborated with world knowledge to form a network according to the relationships existing

among them, which are specified by the text, and the interrelationships existing in the long-term memory. These symbolic rules are weak, general rules rather than strong, smart rules. As a result of these text- and knowledge-based construction processes, a coherence matrix C is defined that specifies link strength among the propositional nodes derived from the text and from knowledge, and an activation vector A is defined that specifies an activation value for each node in the coherence matrix. An important feature of the construction phase is that the resulting representation contains associations and inferences that are irrelevant to a given context as well as relevant ones.

In the integration phase, activation spreads through the network via the links among the nodes until it stabilizes. Following connectionist principles of constraint satisfaction (e.g., Rumelhart & McClelland, 1986), the integration computation involves the post multiplication of the coherence matrix C by the activation vector A until the change in mean activation value after a multiplication is less than some criterion value. In the current model, this value is set to 0.001. All activations are renormalized after each postmultiplication to keep them from growing out of bounds. The renormalization is done by setting all negative activation values to 0 and dividing all activation values by the maximum activation value so that the highest activation will always have a value of 1 (Kintsch, 1992). By means of this integration process, the CI model achieves what many other theories do with complex, powerful control process rules: It assures that the inferences and knowledge elaborations which have been generated will be contextually appropriate because the integration process filters out the contextually inappropriate inferences and elaborations that had been generated in the initial phase.

Current Implementation

In addition to the general properties sketched above, the current model further specifies memory structure, retrieval mechanism, and construction-integration mechanism. The model is implemented in Common LISP.¹

Memory Structure

The memory structure consists of static long-term memory and dynamic working memory. In the long-term memory resides the knowledge base which is a semantic network of events and concepts. We will describe the knowledge base shortly. The working memory is the model's workbench where the model constructs a representation of the text that it processes as a network of propositional nodes. The working memory incorporates the notion of long-term working memory proposed by Ericsson and Kintsch (1995). The long-term working memory is a portion of long-term memory which is kept highly accessible to maximize the efficiency of skilled performances. Information in long-term working memory is stored in stable form, but reliable access to it may be maintained only temporarily by means of retrieval cues in short-term working memory. The representations that the model constructs in the working memory consist of the propositional nodes derived from the text and those retrieved from the long-term memory. At each processing step, only a limited number of nodes are in the focus of attention and serve as retrieval cues. They constitute short-term working memory. The rest of the network remains available during the process and functions as long-term working memory.

Knowledge Base

The knowledge base contains information about objects, events, and states. A portion of the knowledge base is shown in Figure 4. The representation scheme is

¹The source code is available upon request.

most directly influenced by the notion of "event concept coherence (ECC)" proposed by Alterman (1985; Alterman & Bookman, 1992). Two event descriptions in a text are said to be concept coherent if the positions of the concepts they invoke in the underlying semantic network are connected to one another within a given span of the network subject to constraints on the connecting graph (Alterman & Bookman, 1992). Each node in the network represents an event or state of an object, and is expressed as a proposition. Events or states are linked to each other by a set of relations to form a network. For example, events/states 'take' and 'have' are related in the sense that taking something is an antecedent for having it, or equivalently, having something is a consequence of taking it.

Insert Figure 4 about here.

There are several types of links for connecting events/states in the semantic network (Alterman, 1985). We will only discuss those that are most important to the discussion of the simulations that follow. For details of other link types, refer to Morishima (1996). Of the link types the model recognizes, temporal connection types are of major importance. They are used to chain together typically co-occurring events and states. There are four temporal link types: antecedent (ANTE), predecessor (PRED), consequence (CONSEQ), and sequel (SEQUEL). Antecedent and predecessor events occur before an event while consequence and sequel events come after an event. The temporal connections are also divided into necessary and plausible classes of connections. Antecedent and consequence events are necessary whereas predecessor and sequel events are plausible. For example, LOOK-FOR is an antecedent of FIND; namely, to FIND something one must usually have LOOKed FOR it. On the other hand, OPEN is a predecessor of DRINK; sometimes but not

always before DRINKing it is first necessary to OPEN the container. The model treats necessary and plausible links in different ways. Instruments are connected to events (actions) via the instrument (INSTR) link type. For example, to express 'pounding something with a hammer', (POUND, ?AGT, ?OBJ) is connected to (USE, ?AGT, HAMMER) by the instrument (INSTR) link, where ?AGT and ?OBJ denotes an unspecified agent and object. (see also Figure 4).

Connection strengths for each link are free parameters. They could be estimated based on some empirical data. In the present study, however, the main objective for the modeling work is to provide qualitative simulations for the experimental results rather than quantitative predictions, and thus a priori values are used and then adjusted if necessary to produce desirable outcomes.

Simulation Walkthrough

We will walk through a sample simulation to illustrate the retrieval and construction-integration processes. The following text is used for illustration:

3.3. John took the hammer out of the garage. He pounded the boards together in the afternoon.

In this study, parsing processes are excluded from consideration. Analysis of the text begins with propositional representations which are assumed to be constructed by the parsing component of the system (Kintsch, 1974). The text is analyzed into the following propositions:

P1 (TAKE , JOHN, HAMMER)

P2 (LOC, PROP P1, GARAGE)

P3 (POUND, JOHN, BOARD)

P4 (TIME, P3, AFTERNOON)

The propositions, P1 and P2, come from the first sentence, and P3 and P4 from the second sentence. It is assumed that the basic unit of processing is a sentence. Hence, P1 and P2 are processed together first while P3 and P4 are processed together in a later processing period.

STEP 1:

The elaboration process was performed after processing all the propositions from the text since it is legitimate to assume that the comprehender does not engage in extensive elaboration of a text until all sentences are read especially when the length of the text is very short and in an experimental setting where the comprehender usually does not have a motivation for elaboration. The initial propositional network is constructed with the propositions P1 and P2, which are connected to each other with a bi-directional link with the strength of 1.0. The propositions P3 and P4 are processed in the same manner as P1 and P2 are processed. In the network shown in Figure 5, the black nodes and the connections among them are constructed here.

Now the elaboration process begins; the text propositions retrieve pieces of knowledge related to them from LTM. In order for a proposition to serve as a retrieval cue for elaboration, it needs to be a node newly added to the network and must have the activation value equal to or greater than the threshold. In this example, the threshold is set to 0.4. The four text propositions in the current network qualify because they are new propositions and their activation values are 1.0. These nodes retrieve the long-term memory propositions P5 (USE, JOHN, HAMMER), P7 (TOOL, HAMMER, POUND)², P8 (FIND, JOHN, HAMMER), and P9 (HAVE, JOHN, HAMMER), which are connected to their retrieval cue propositions

²Not only propositions but also arguments within the propositions can serve as retrieval cues. In this case, the object HAMMER in P1 retrieves (TOOL, HAMMER, POUND), which represents that HAMMER is a TOOL for POUNDing.

with the link types specified in the LTM, and assigned the activation value of 0.0 and the self-strength of 1.0. For example, P9 (HAVE, JOHN, HAMMER) is connected to P1 (TAKE (AGT JOHN) (OBJ HAMMER)) with the link type of CONSEQ, and it denotes that John took a hammer and as a result he had it. Note that P3 (POUND, JOHN, BOARD) has retrieved P5 (USE, JOHN, HAMMER) and P6 (NOT (USE, JOHN, HAMMER))³ and they are connected to each other by an inhibitory link. The negation proposition P6 thus competes with P5. This means that given (POUND, JOHN, BOARD), it is probable that he used a hammer, but he may not have. However, the connection strength between (POUND, JOHN, BOARD) and (USE, JOHN, HAMMER) is twice as great as that between (POUND, JOHN, BOARD) and (NOT (USE, JOHN, HAMMER)). This indicates that there is a bias toward the affirmative proposition; namely, it is more likely that one assumes that John used a hammer than one assumes otherwise. Then the network goes through the integration phase. In the construction phase of the next step, those long-term memorypropositions that have been added to the network in Step 1 are potential retrieval cues. In fact, propositions P7, P8, and P9 have gotten the activation values greater than the threshold of 0.4 as a result of integration, and thus serve as retrieval cues (marked by a thick circle in Figure 5).

STEP 2:

The system further elaborates the representation with P7 (TOOL, HAMMER, POUND), P8 (FIND, JOHN, HAMMER), and P9 (HAVE, HAMMER) as retrieval cues in Step 2. P7 (TOOL, HAMMER, POUND) connects itself to P5 (USE, JOHN, HAMMER) when it retrieves (USE, JOHN, HAMMER). Note that connections between P1 and P8, and P1 and P9 are now bi-directional. This is because P8 (FIND, JOHN, HAMMER) and P9 (HAVE, HAMMER) retrieve (TAKE, JOHN, HAMMER)

³The knowledge base does not contain negation propositions. A negation proposition such as this one is generated during the construction phase.

and make connections to it. As a result of the construction process in Step 2, the network as shown in Figure 5 is built in the working memory. Again, the network is submitted to the integration phase. When the integration process is done, the system checks the activation values of those newly added propositions (i.e., P10 through P15), and finds that none of them has gained an activation value equal to or greater than the threshold. Thus, no more knowledge is retrieved from LTM, and the processing is completed.

Insert Figure 5 about here.

Simulation Results

Several simulation runs were conducted with different texts. The main goal of these simulations is to provide an adequate account for the experimental data of the present study. For this purpose, simulations were intended to be qualitative in that they showed the major trends in the experimental results, and extensive parameter estimation was not attempted.⁴ As in the experiments, texts were constructed by combining each of the four context sentences and the action sentence as shown below:

- 3.4a. John broke the hammer in the garage. (Contradictory)
- 3.4b. John looked for the hammer in the garage. (Weak)
- 3.4c. John found the hammer in the garage. (Moderate)
- 3.4d. John took the hammer out of the garage. (Strong)
- 3.5. He pounded the board in the afternoon. (Action)

⁴Of course, there are drawbacks of qualitative simulations. For example, there is no statistical evaluation, and thus it leaves room for disagreement about validity of a model (Kintsch, 1992).

Simulations of Experiments 1 and 2

Experiments 1 and 2 showed that instrument inference was drawn only when the context strongly supports the inference. We assume that in these experiments the participants did not have any specific goals and hence were not motivated to adopt an elaborative reading strategy. In the simulation, this assumption is realized by setting a high activation threshold (0.4). The simulation run with the strongly related context has been shown in the above walkthrough (see Figure 5). The contradictory text and the moderately related text were also submitted to simulation in the same manner as the walkthrough.

Figure 6 shows the time course of activation of the instrument proposition for the strong context (i.e., the TAKE-HAMMER text) and moderate context (i.e., the FIND-HAMMER text). For the first ten processing cycles (Step 1) the activation patterns are identical for both the strong and moderate contexts. This is because in Step 1 of processing, for both contexts, the inference proposition is retrieved by the action proposition (POUND, JOHN, BOARD) and receives activation from it. However, in Step 2, for the strong context, the propositions (HAVE, JOHN, HAMMER) and (TOOL, HAMMER, POUND) make connections to the USE-HAMMER proposition and send activation to it, and as a result a higher activation is achieved (see Figure 5).

Insert Figure 6 about here.

On the other hand, for the moderate context, while the TOOL proposition sends some activation to the USE-HAMMER proposition, (HAVE, JOHN, HAMMER) does not make a connection to the USE-HAMMER proposition because it does not reach the activation level high enough to be a retrieval cue (see Figure 7).

As a result, the activation steadily decreases. When the networks were settled after 24 processing cycles, the activation of the inference for the strong context was higher than that for the moderate context. In Experiment 2, the results showed the mean reaction time for the strong context was faster than that for the moderate context. Thus, the model has successfully simulated the results.

Insert Figure 7 about here.

The simulation run with the contradictory context exhibited a different pattern of behaviors. The activation of negation proposition (NOT (USE, JOHN, HAMMER)) shot up quite rapidly and kept a high value throughout the processing, whereas (USE, JOHN, HAMMER) received some activation initially but lost it quickly. This is because the NOT-USE-HAMMER proposition is connected to the BREAK-HAMMER proposition and receives activation and wins the competition against the USE-HAMMER proposition, which receives only much smaller amount of activation from the POUND-BOARD proposition due to the smaller link strength for the INSTRUMENT link. Thus, the simulation predicted that for the contradictory condition, the instrument inference would not be drawn. Indeed, Experiment 1 found the instrument priming for the strongly related context but not for the contradictory context.

Simulations of Experiment 3

In Experiment 3, the participants were trained to attend to the use of the instrument by answering comprehension questions about instruments, and instrument priming was observed for both the moderate and strong contexts. This result showed that when the comprehender was engaged in a deeper processing to

make more elaboration, inferences were made as long as the context is compatible with such an inference.

To simulate these results, the activation threshold was lowered to 0.1 in the subsequent set of simulations. This made the system pay attention to and make use of those propositional nodes with lower activation, which did not function as retrieval cues in the earlier simulations. The same sets of texts were used for these new simulations. Figure 8 shows the comparison of the final activation of the inference between the high threshold and the low threshold simulation runs for each text. For the strong context, the inference achieved high activation for both thresholds. For the weaker contexts, where the activation of the inference was low for the high threshold, higher level of activation resulted when the threshold was lowered. For the contradictory context, the change in the threshold did not have any effect; namely, the inference node was not activated at all in either case. These outcomes are consistent with the experimental results.

Insert Figure 8 about here.

A closer look at the model's comprehension process for the FIND-HAMMER text reveals more details about how these results came about. The activation patterns for the inference proposition in both simulation runs look identical during the first two steps of the processing. However, in Step 3, the activation increased because the HAVE-HAMMER proposition, which had been added to the network in Step 2, made a connection to the USE-HAMMER proposition and sent activation to it (compare Figures 7 and 9). This indicates that the integration of the proposition (HAVE, JOHN, HAMMER) into the representation is crucial in activating the inference proposition because it functions as the major source of facilitation for the

inference. In other words, as discussed in the experiment section, the presence of the HAVE-HAMMER proposition serves as a precondition for the inference to be drawn.

Insert Figure 9 about here.

4. SUMMARY AND CONCLUSIONS

In the present study, we hypothesized that the structure of a situation model influences elaborative instrument inferences during comprehension. The experimental results give support to the hypothesis. The results showed that under normal reading without any specific reading goals, the degree of elaboration involved in construction of a situation model is not extensive. However, if the comprehender is motivated to elaborate and retrieve more knowledge during the process, the inference is computed even in the context that weakly supports the inference. The CI model successfully displayed the behaviors that are qualitatively in agreement with the experimental results reported in the previous section. The model provides an explicit account for the inference computation processes and makes predictions about such processes.

In conclusion, On-line inference processes, particularly elaborative inferences, are influenced by the trade-off between the demand for constructing a rich situation model and the limited capacity of the human cognitive system. On the one hand, the comprehender tries to activate and integrate as much world knowledge as possible to construct a situation model as much elaborated by world knowledge as possible. On the other hand, due to the limited capacity, the comprehender can allow only a limited amount of resources for the on-line inference processes. As a

consequence, elaborative inferences are strongly activated and encoded on-line only when a text contains information specific enough to lead to information stored in the episodic text memory or the knowledge base. Otherwise, the encoding of such inferences would not be made or, even if they are computed, would not be strong enough to be detected experimentally.

REFERENCES

- Alterman, R. (1985). A dictionary based on concept coherence. *Artificial Intelligence, 25*, 153-186.
- Alterman, R., & Bookman, L. A. (1992). Reasoning about a semantic memory encoding of the connectivity of events. *Cognitive Science, 16*, 205-232.
- Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles – a problem of polysemy. *Cognitive Psychology, 7*, 167-180.
- Collins, A., Brown, J. S., & Larkins, K. M. (1980). Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Corbett, A. (1984). Prenominal adjectives and the disambiguation of anaphoric nouns. *Journal of Verbal Learning and Verbal Behavior, 23*, 683-695.
- Corbett, A. T., & Chang, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition, 11*, 283-294.
- Corbett, A. T., & Doshier, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior, 17*, 479-491.
- 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.

Dosher, B. A., & Corbett, A. (1982). Instrument inferences and verb schemata.

Memory & Cognition, 10, 531-539.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological*

Review, 102 (2), 211-245.

Fincher-Kiefer, R. (1995). Relative inhibition following the encoding of bridging

and predictive inferences. *Journal of Experimental Psychology: Learning,*

Memory, and Cognition, 21 (4), 981-995.

Gernsbacher, M. A. (1989). Mechanisms that improve referential access. *Cognition*,

32, 99-156.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to

foregrounding during text comprehension. *Journal of Memory and Language*,

26, 69-83.

Graesser, A. C., & Kreuz, R. J. (1993). A theory of inference generation during text

comprehension. *Discourse Processes*, 16, 145-160.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a

process in comprehension. *Journal of Verbal Learning and Verbal Behavior*,

13, 512-521.

Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory for tacit

implications of sentences. *Journal of Experimental Psychology*, 98, 203-205.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University

Press.

Keefe, D. E., & McDaniel, M. A. (1993). The time course and durability of predictive

inferences. *Journal of Memory and Language*, 32, 446-463.

- Keenan, J. M., & Jennings, T. M. (1995). Priming of inference concepts in the construction-integration model. In C. A. Weaver, III, S. Mannes, & C. R. Fletcher (eds.), *Discourse Comprehension: Essays in honor of Walter Kintsch*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 93 (2), 163-182.
- Kintsch, W. (1992). A cognitive architecture for comprehension. In H.L. Pick, Jr., P. van den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues*. Washington, D.C.: American Psychological Association.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kintsch, W., & Vipond, D. (1978). Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Proceedings of the conference on memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kintsch, W., & Welsch, D. M. (1991). The construction-integration model: A framework for studying memory for text. In W. E. Hockley & S. Lewandowsky (Eds.), *Perspectives of memory research*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lesgold, A. M., Roth, S. F., & Curtis, M. E. (1979). Foregrounding effects in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior*, 18, 291-308.
- Lucas, M. M., Tanenhaus, M. K., & Carlson, G. N. (1990). Levels of representation in the interpretation of anaphoric reference and instrument inference. *Memory & Cognition*, 18, 611-631.

- Maurer, G., Tanenhaus, M. K., & Carlson, G. N. (1995). Implicit arguments in sentence processing. *Journal of Memory and Language*, 34, 357-382.
- McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*, 19, 668-682.
- McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrument inference. *Journal of Verbal Learning and Verbal Behavior*, 20, 671-682.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12 (1), 82-91.
- McKoon, G., & Ratcliff, R. (1989). Assessing the occurrence of elaborative inference with recognition: compatibility checking vs compound cue theory. *Journal of Memory and Language*, 28, 547-563.
- Morishima, Y. (1996). *Effects of Discourse Context on On-Line Inference Computation*. Unpublished doctoral dissertation, University of Colorado.
- Murray, J. D., Klin, C., M., & Myers, J. L. (1993). Forward inferences in narrative text. *Journal of Memory and Language*, 32, 464-473.
- Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, 26, 453-465.
- Newell, A. (1973). Production system: Models of control structure. In W. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- Norvig, P. (1989). Marker passing as a weak method for text inferencing. *Cognitive Science*, 13, 569-620.

- O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K (1988). Elaborative inferences during reading: do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14 (3), 410-420.
- Paris, S. G., & Lindauer, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology*, 8, 217-227.
- Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, 27, 399-415.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing. Explorations in the microstructure of cognition. Vol. 1: Foundations.* Cambridge, MA: MIT Press.
- Sanford, A. J., & Garrod, S. C. (1980). *Understanding written language.* New York: Wiley.
- Singer, M. (1979). Processes of inference during sentence encoding. *Memory & Cognition*, 7 (3), 192-200.
- Singer, M. (1980). The role of case-filling inferences in the coherence of brief passages. *Discourse Processes*, 3, 185-201.
- Singer, M., & Ferreira, F. (1983). Inferring consequences in story comprehension. *Journal of Verbal learning and Verbal Behavior*, 22, 437-448.
- van Meter, P., & Pressley, M. (1994). Encoding of instruments when 10- to 14-year-olds process isolated instrument-implicit sentences: More evidence of improved encoding during childhood resulting from elaborative instructions. *Journal of Educational Psychology*, 86 (3), 402-412.

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

Whitney, P., Ritchie, B. G., & Crane, R. S. (1992). The effect of foregrounding on readers' use of predictive inferences. *Memory & Cognition*, 20 (4), 424-432.

Table 1. Mean Reaction Times (in msec), Standard Deviations, and Percentages of Error for Experiment 1.

SOA	Context								
	Unrelated			Contradictory			Related		
	RT	SD	Err(%)	RT	SD	Err(%)	RT	SD	Err(%)
250 msec	1111	209.5	5	1101	214.8	4	1077	194.2	6
500 msec	1082	207.6	4	1089	227.6	3	1042	196.0	3

Table 2. Mean Reading Times (in msec) and Standard Deviations for Experiment 1.

SOA	Context					
	Unrelated		Contradictory		Related	
	RT	SD	RT	SD	RT	SD
250 msec	1648	603.5	1650	461.3	1438	439.9
500 msec	1747	520.6	1805	550.0	1539	422.2

Table 3. Mean Reaction Times for Lexical Decision and Mean Reading Times of Action Sentences for Experiment 2.

Context	RT for Lexical Decision	Reading Time
Unrelated	844 ^a (188.7) ^b	1578 (555.0)
Weak	818 (177.3)	1336 (423.1)
Moderate	814 (155.8)	1332 (510.6)
Strong	755 (135.5)	1270 (444.0)

^aThe values are all in msec. ^bThe values in parentheses are standard deviations.

Table 4. Mean Reaction Times for Lexical Decision, Mean Reading Time of Action Sentences, and Mean Response Time for Comprehension Questions for Experiment 3.

Context	RT for Lexical Decision	Reading Time	RT for Question
Contradictory	765 ^a (210.7) ^b	1495 (492.5)	808 (419.1)
Moderate	723 (187.7)	1399 (486.9)	859 (421.9)
Strong	712 (208.3)	1402 (489.1)	786 (373.4)

^aThe values are all in msec. ^bThe values in parentheses are standard deviations.

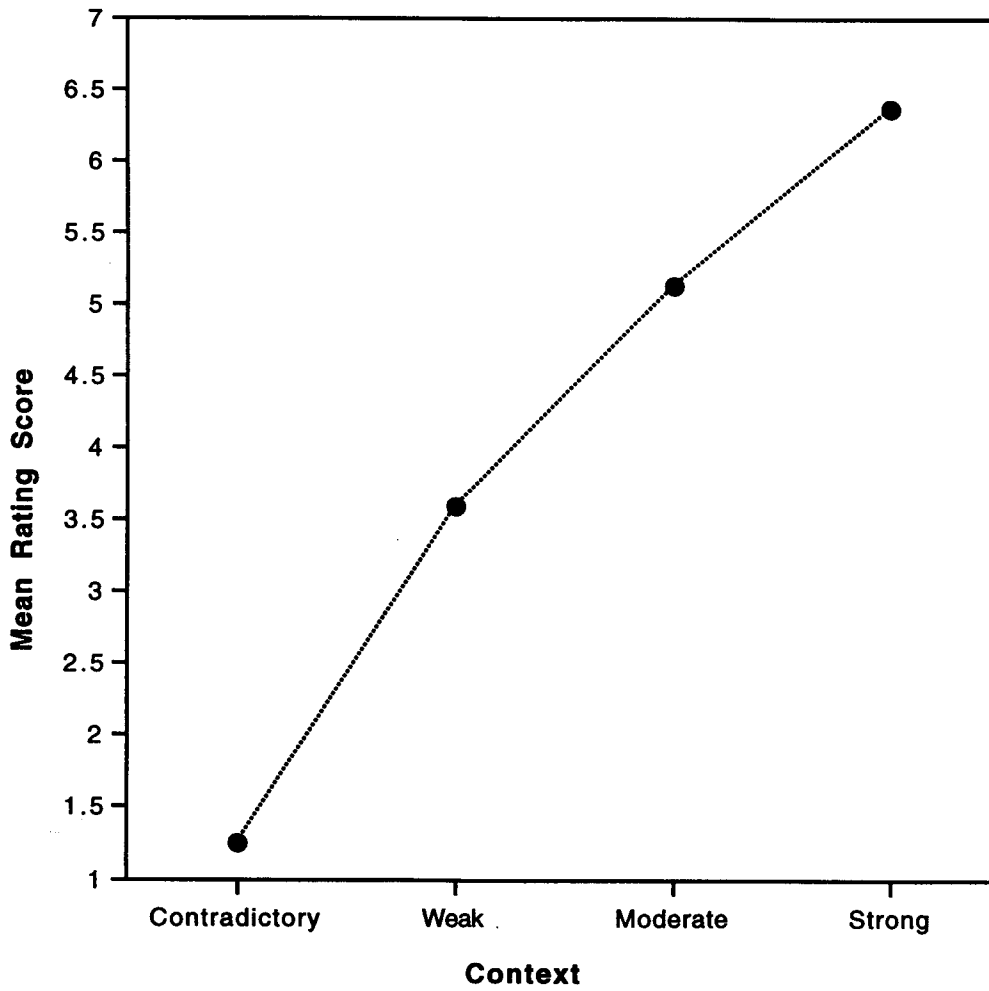


Figure 1. Mean rating scores of the experimental texts as a function of the level of context.

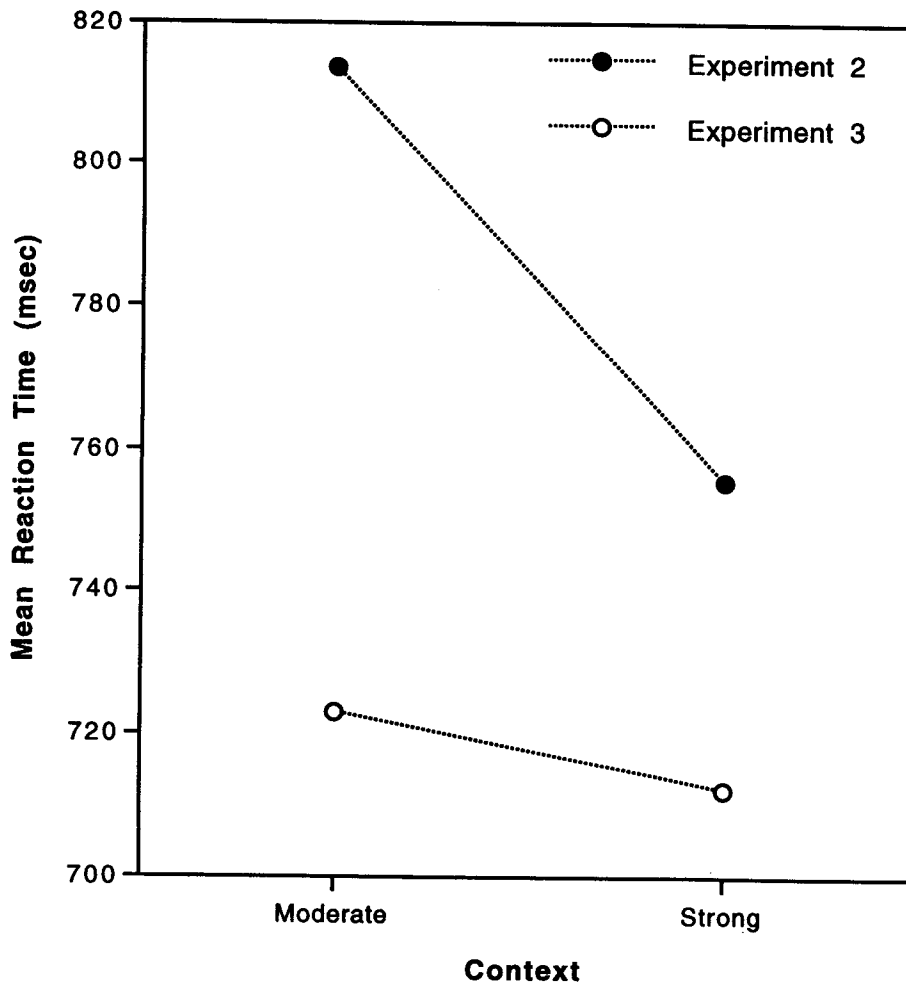


Figure 2. Comparison of reaction times between Experiment 2 and Experiment 3.

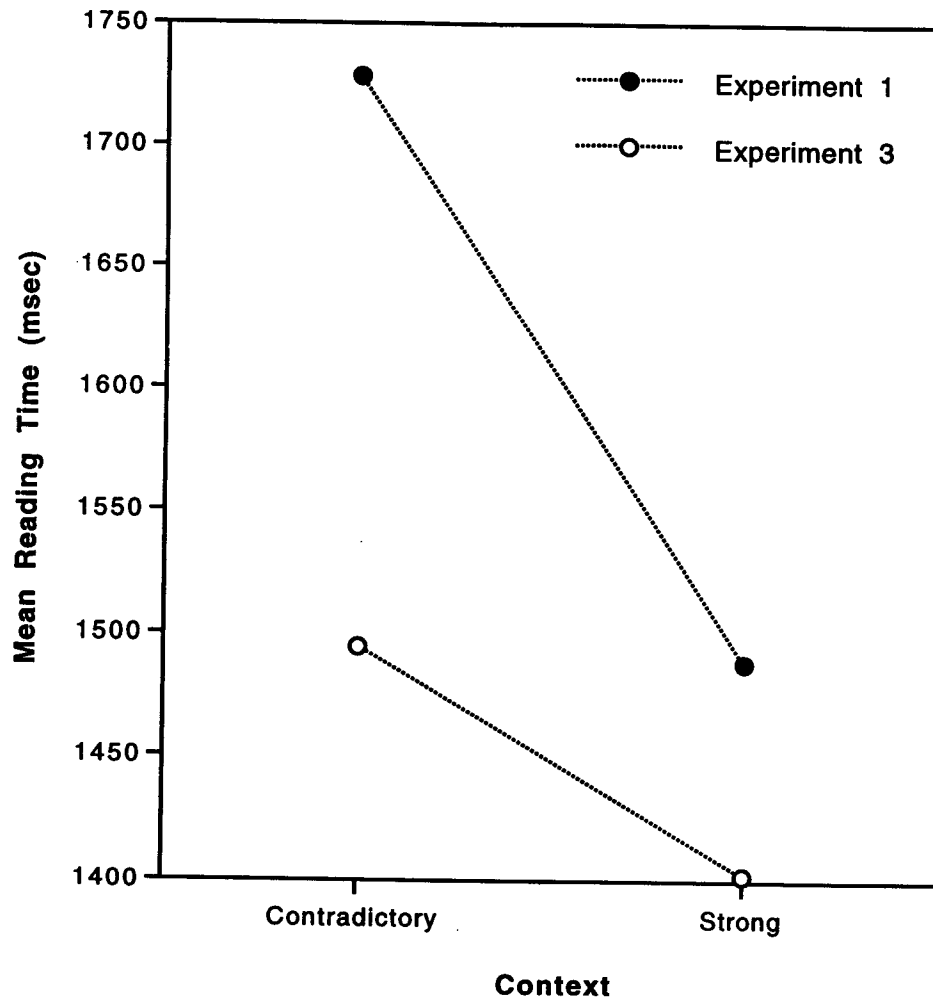


Figure 3. Comparison of reading times between Experiment 1 and Experiment 3.

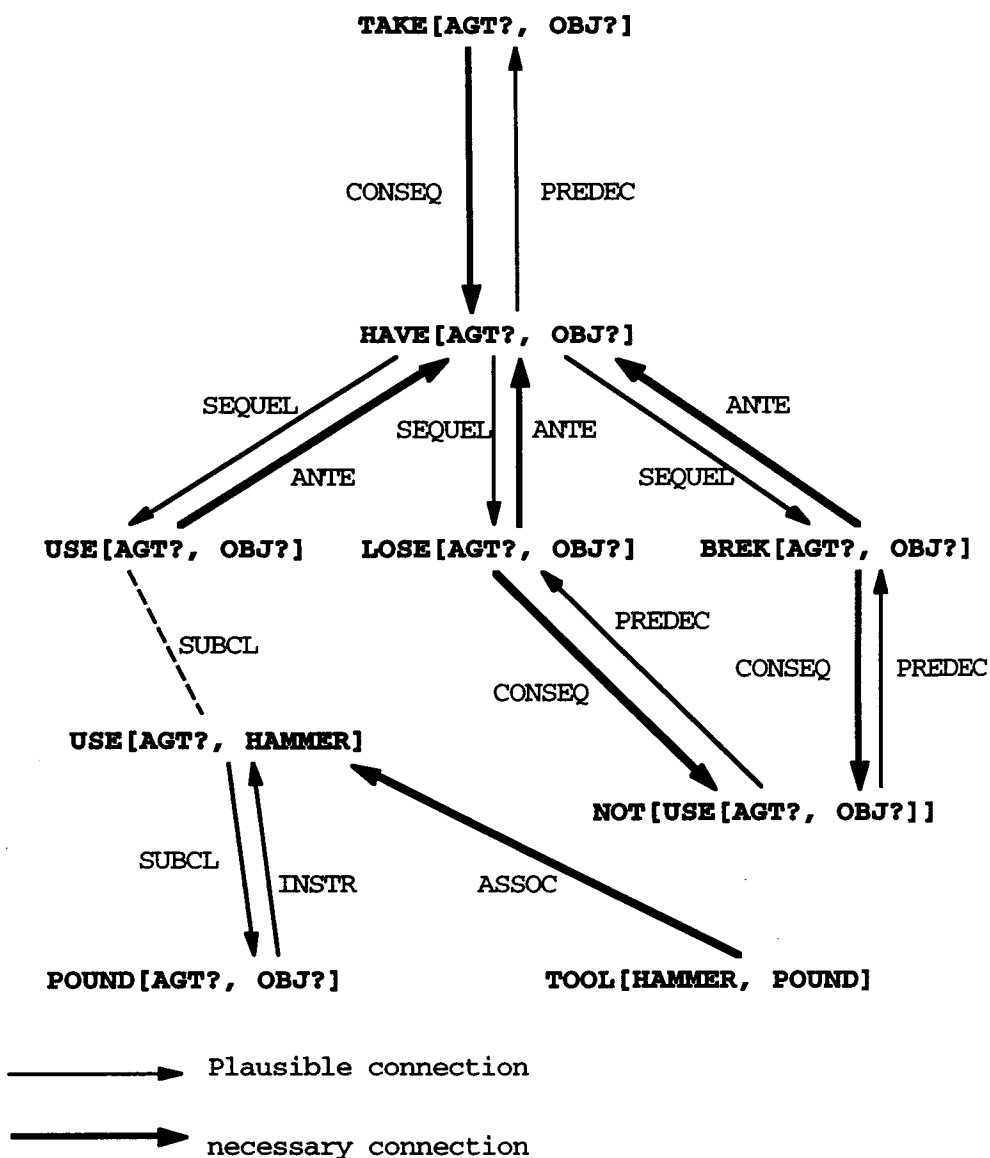


Figure 4. A portion of long-term memory semantic network.

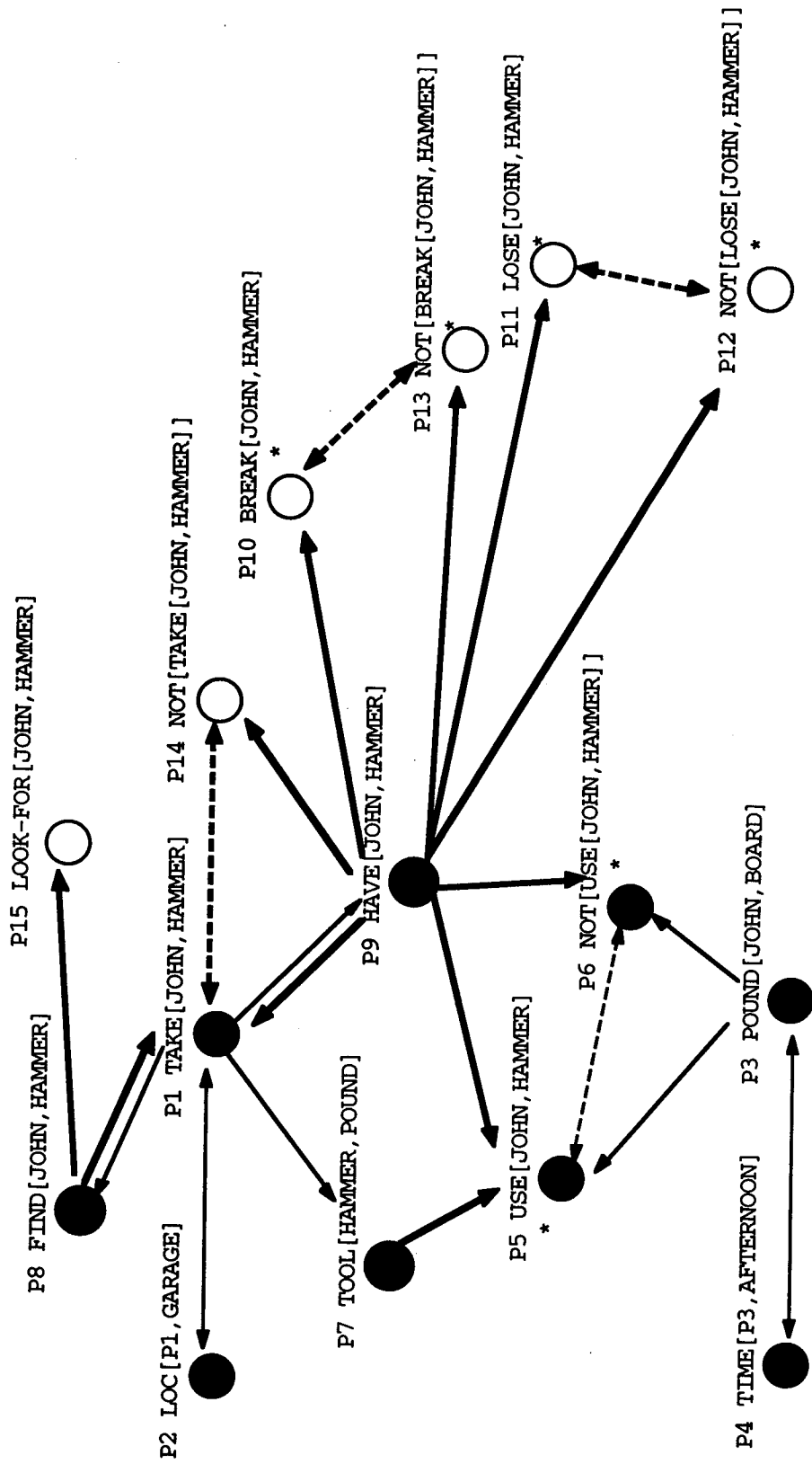


Figure 5. The propositional network for the text: *John took the hammer out of the garage. He pounded the boards together in the afternoon.* The propositional nodes with thick borders are retrieval cues for the next processing step. The broken arrow indicates an bidirectional inhibitory link. The nodes with asterisks are fully connected to each other by inhibitory links.

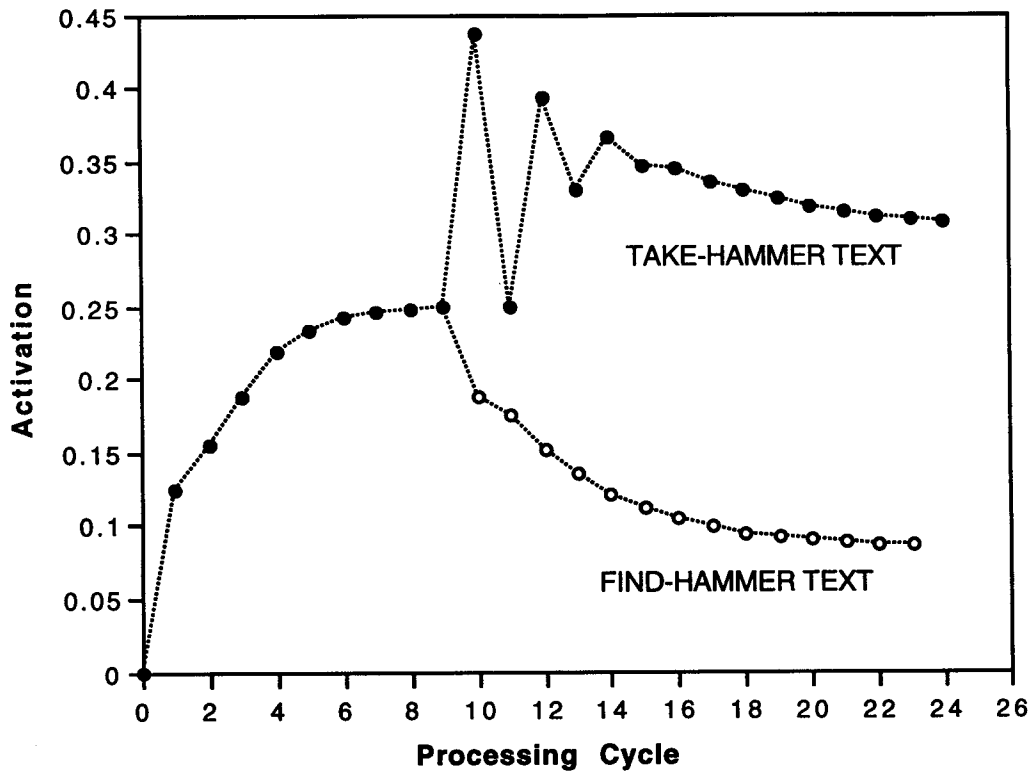


Figure 6. Time course of activation of the inference proposition USE-HAMMER for the TAKE-HAMMER text and the FIND-HAMMER text with a high activation threshold for retrieval.

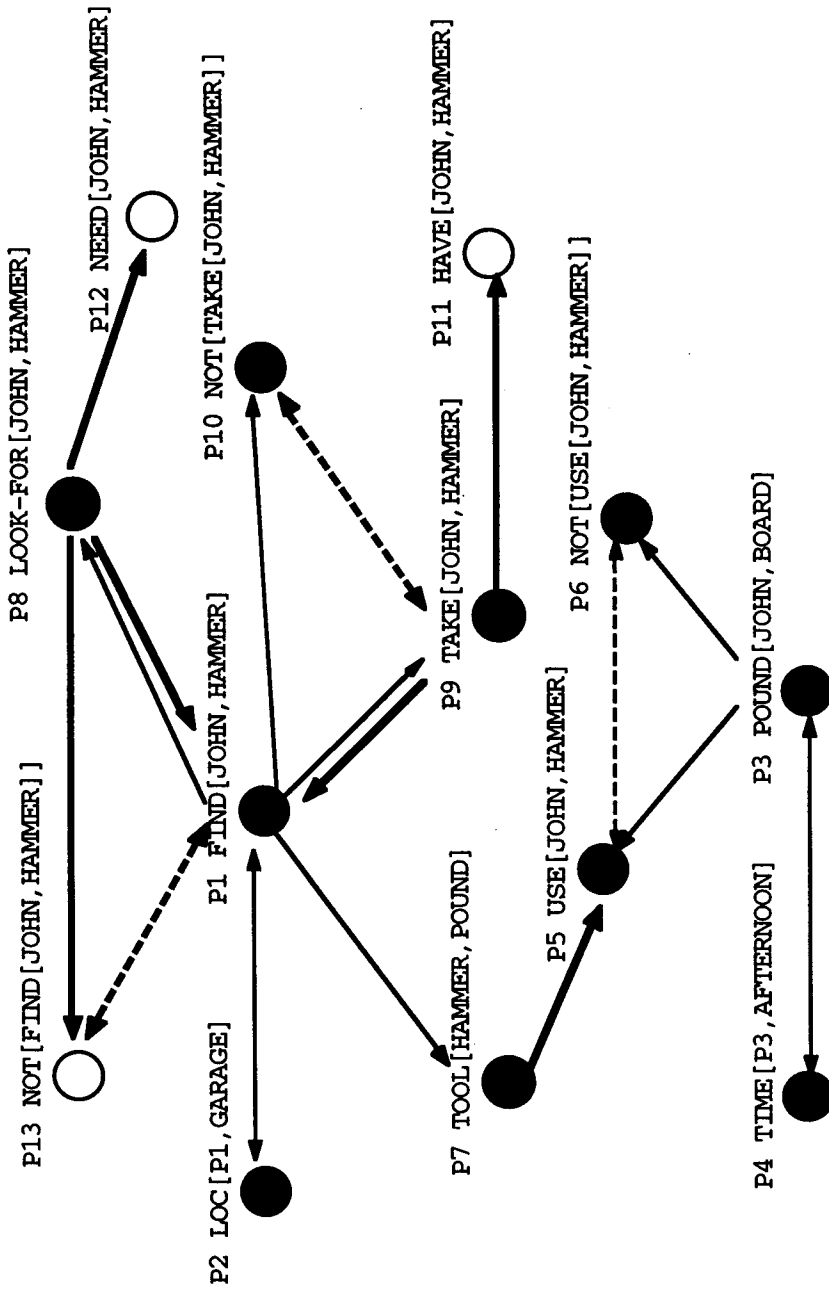


Figure 7. The propositional network for the text: *John found the hammer in the garage. He pounded the boards together in the afternoon.* The propositional nodes with thick borders are retrieval cues for the next processing step. The broken arrow indicates an inhibitory link.

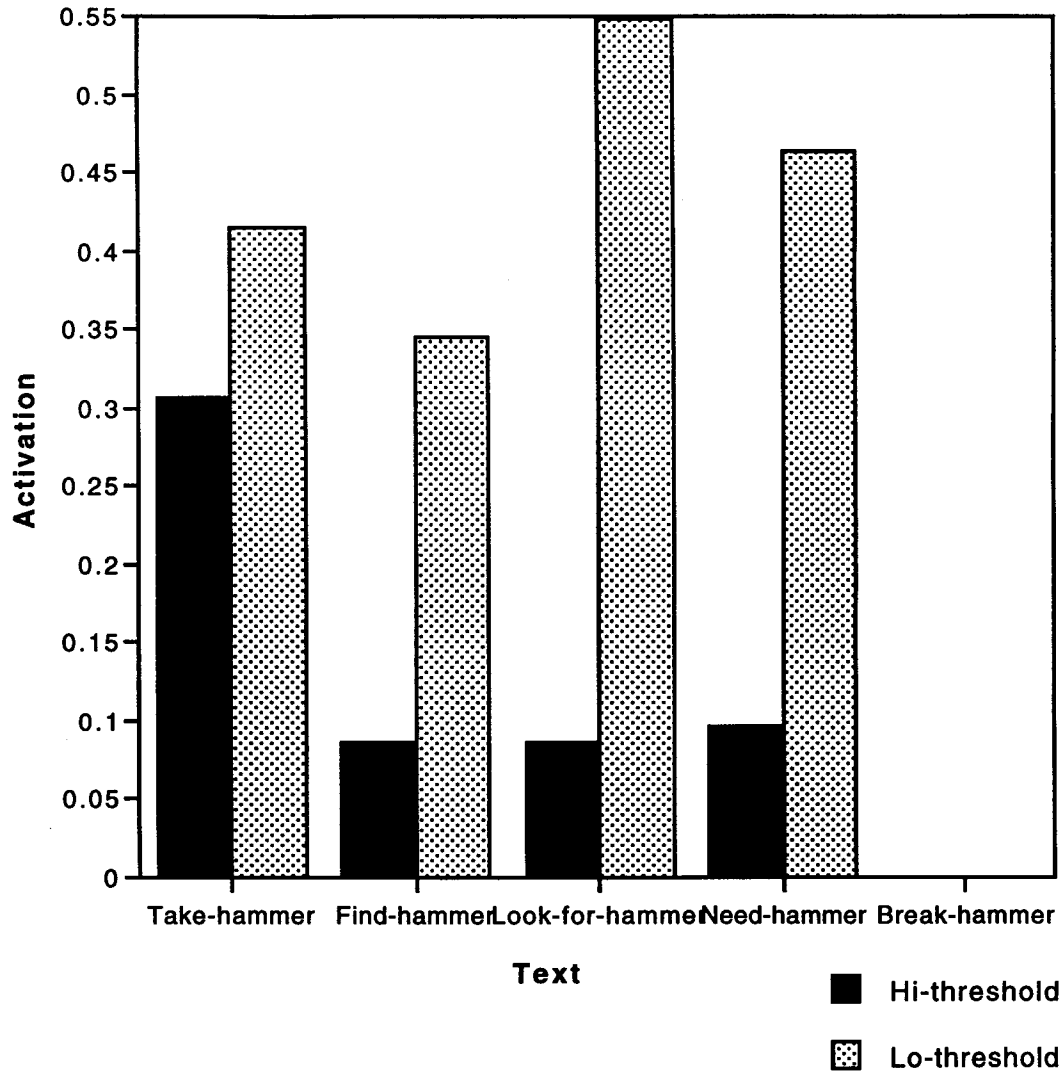


Figure 8. Final activation of the USE-HAMMER proposition for different texts.

