Frame-shifting and Sentential Integration

Seana Coulson
Department of Psychology
University of Arizona
Marta Kutas
Department of Cognitive Science
University of California, San Diego

Technical Report 98.03

Department of Cognitive Science
University of California, San Diego
La Jolla, CA 92093-0515

October 16, 1998

, 1998. The reference for this document is: Technical Report CogSci.UCSD-98.03, October, 1998, Department of Cognitive Science, UCSD, San Diego, CA, 92093-0515.

Frame-shifting and Sentential Integration

Seana Coulson
Department of Psychology
University of Arizona
Marta Kutas
Department of Cognitive Science
University of California, San Diego

Abstract

We consider different ways in which understanding a single word can affect the overall construal of the discourse event. One such process is *frame-shifting*, semantic reanalysis in which elements of the existing message-level representation are mapped into a new frame. This view predicts that scenarios which occasion frame-shifting present a challenge to the processor different from that presented by lexical violations consistent with the currently active frame. To test this hypothesis, we manipulated the relationship between sentence final words and their preceding contexts, and compared reading times for words which triggered frame-shifting, with those for equally unexpected but frame-consistent words. In both humorous and nonhumorous sentences, participants spent longer reading words which triggered frame-shifting than frame-consistent controls. The pattern of results suggests that the message-level representation involves hierarchically organized attribute-value structures, and includes causal and relational information. Traditional frame-based comprehension systems, however, cannot fully account for the dynamically adaptive nature of frame-shifting. An alternative is mentioned in brief.

1 Introduction

It often seems that life would be easier if people just said what they meant. For example, depending on the speakers and circumstances, "That's a nice outfit," might be an observation, a compliment, an insult, or a request to change clothes. While it is possible to assign abstract meanings to words and sentences, the meanings they assume in particular utterances can be quite different. Moreover, perhaps as an acknowledgment of the potential gap between meanings in and out of context, researchers interested in the comprehension of what is *said* have labored somewhat independently of those interested in what is *meant* (McKoon & Ratcliff, 1998). For instance, researchers in text processing have been concerned with the end-product of comprehension (see Zwaan & Radvansky, 1998 for review) while those in sentence processing have focussed more on how the message-level representation influences earlier processes. A major (in fact, one might say, *the* major) focus of sentence processing research in the 90s has been whether contextual factors influence the initial assignment of a syntactic representation, and how context affects syntactic reanalysis (e.g. Frazier, 1995; Spivey-Knowlton & Tannenhaus, 1994). Similarly, a key issue in research on word recognition has been whether context affects the initial choice of a word's meaning (see Simpson, 1994 for review).

In short, mainstream research has primarily concerned how context influences the processing of individual words, rather than how words contribute to the development of the representation at the message-level (Sharkey & Mitchell, 1988). Pursuing a slightly different approach to the relationship between lexical and contextual processing, we consider different ways in which understanding a single word can affect the overall construal of the discourse event. Striving for a model which can explain both what is said and what is meant, we entertain the hypothesis that conceptual operations invoked to construct the message-level representation might similarly be involved in lexical processing. We begin by sketching a general framework for thinking about the interaction of sentence and text processing, and investigate how a particular operation we call *frame-shifting* affects meaning construction at the message-level.

1.1 Frames

Perhaps the most central characteristic of meanings is their variability across contexts. As noted above, the interpretation of an utterance is modulated by the speaker, the intended audience, the local surroundings, and world knowledge about the scenario in question (e.g., Clark, 1983; Sanford & Garrod, 1980; Travis, 1981). This variability in meaning is sustained, at least in part, by speakers' ability to recruit background knowledge to construct meaning for utterances. For example, contrast the meaning of 'bouncing' in the following passages (adapted from Sanford & Garrod, 1981):

(1a) Jaimie came bouncing down the stairs.
(1b) Paul ran over to kiss her.
(2a) Jaimie came bouncing down the stairs.
(2b) Paul rushed off to get the doctor.

While in (2a) Jaimie tumbles dangerously down the stairs, in (1a) she walks down with excited, springing steps. To understand the difference in these two 'bounces', it is necessary to invoke background knowledge to connect Paul's actions in the (b) sentences to Jaimie's actions in the (a) sentences. The ease with which people seem to make these sorts of connections suggests that processes of meaning construction routinely appeal to this sort of causal and relational information (Rumelhart, 1980; Sanford & Garrod, 1981). In fact, Lange & Dyer (1989) argue that natural language processing requires a capacity for dynamic inferencing based on general knowledge rules represented as *frames*.

We suggest that language cues the retrieval of frames from long-term memory, and that these frames (or data structures) are then exploited in the construction of mental models of the message-level representation. As used here, frames are representations with slot/filler structure, default values, and weak constraints on the type of fillers for a given slot (reviewed in Barsalou, 1992; Rumelhart & Ortony, 1977). Frames contain causal and relational information, are organized hierarchically so as to allow recursive embedding of frames within frames, and can be used to represent knowledge about a wide variety of objects, actions, and events (Fillmore, 1968; Minsky, 1975; Sanford & Garrod, 1980; Schank & Abelson, 1977). Invented somewhat independently by researchers in the fields of linguistics, psycholinguistics, artificial intelligence, and natural language processing, frame-type data structures are often invoked to account for inferential aspects of language comprehension (Fillmore, 1982).

The basic idea is that both linguistic and nonlinguistic cues evoke frames which can be bound to contextually available elements. For example, the restaurant script might be activated to understand a story about a dinner date. Comprehension would proceed by binding slots in the script (such as the Waitress slot) to particular elements available from the discourse. Other slots in the frame would imply the existence of other elements (such as the Maitre d'). Because slots can be thought of as implicit hypotheses about entities in the discourse, part of their computational utility is to extend the domain of reference to implied entities in the scene. Further, binding elements into slots in a frame affords a way of integrating them into the overall message-level representation of meaning. Besides providing a way of understanding the relationship between new and established elements, binding an element to a slot can help to constrain the overall interpretation of the utterance.

At some level, researchers have appreciated the importance of frame-based knowledge in language comprehension for many years (e.g. Bower, Black & Turner, 1979; Graesser, Woll, Kowalski, & Smith, 1980). For example, Rumelhart (1980) tested the role of frames in story comprehension by asking participants to answer a series of questions after each sentence in a short vignette. Their progressive responses to these questions – for example, what is going on in the story? who are the characters? and why are they behaving as they are? – suggested a process of frame activation and refinement. To allay skeptics' concerns that protocols do not reflect on-line comprehension processes, Rumelhart also collected reading time data for one particular vignette in which protocol analyses suggested that readers dramatically shifted their interpretation during the last sentence of the story. As illustrated below, two versions of the vignette were constructed, one consistent with most people's initial interpretation of the paragraph as being about a woman's pet, and another in which the reader learns that the paragraph is actually about a fur:

Dear little thing. It was nice to feel it again. She had taken it out of its box that afternoon, given it a good brush and rubbed life back into its dim little eyes. Little rogue! Yes, she really felt that way about it. She put it down/on. Little rogue, biting its tail just by her left ankle/ear. When she breathed something gentle seemed to move on her bosom. The day was cool and she was glad she had decided on her little pet/fur.

Rumelhart gave one version of the paragraphs to each of two different groups and compared their reading times for the first 49 words (whose content was identical for both groups), and for the last 38 words (which differed at three key points). He found that even when corrected for faster average reading rates in the "Pet" group, participants in the "Fur" group spent reliably longer on the latter part of the story, presumably because comprehension required reanalysis of their initial "Pet" interpretation. By violating our expectations, Rumelhart's Fur passage underscores the extent to which frames from background knowledge are continuously used to structure expectations and make inferences that go beyond what is immediately present.

Moreover, Rumelhart's finding – that participants dwell longer on the "Fur" passage because it fails to match their "Pet" frame – presents a number of interesting issues. For example, exactly what processing operations occur during the extra time it takes participants to read the "Fur" passage? How and when do participants realize that their original interpretation no longer applies? This finding also highlights an inherent problem with a strictly frame-based approach to comprehension: namely, that frame-based inferences often turn out to be wrong. What process(es) are invoked to deal with this problem? What happens to the information that has already been encoded? How does the reader know what information to maintain, and what information to reject for a reinterpretation?

1.2 Frame-shifting

We call the conceptual reanalysis necessary to understand passages such as the "Fur" vignette *frame-shifting*, because it seems to require reorganizing information in the message-level representation into a new frame. Minsky (1980) notes the prominence of frame-shifting in many kinds of jokes:

> "The element that seems to me most common to all the different kinds of humor is that of unexpected frame-substitution, in which a scene is first described from one viewpoint and then suddenly – typically by a single word – one is made to view all the scene-elements in another, quite different way, (Minsky, 1980: 10)."

In fact, jokes are deliberately constructed to suggest one frame while evoking elements consistent with another. While frame-shifting is not unique to jokes, jokes differ from more 'everyday' examples in the extent to which the need to shift is clearly demarcated. Joke theorists such as Attardo (1990) call this cue the *disjunctor*. For example, in (3) the reader begins by evoking a frame where a busy professional pays an accountant to do his taxes.

(3) I let my accountant do my taxes because it saves time: last spring it saved me ten years.

At the disjunctor "years," however, the reader is forced to go back and reinterpret "time" to evoke a frame where a crooked businessman pays an accountant to conceal his illegal business dealings. The word "time" is called a *connector* because it serves as a bridge between the two frames. Merely knowing that "time" refers to time in prison does not in and of itself explain why the accountant is doing the man's taxes, or how doing so will prevent a prison sentence. A full understanding of (3) requires recruitment of background knowledge about the particular sorts of relationships that can obtain between business people and their accountants. The initial busy-professional interpretation is thus mapped into the crooked-businessman frame.

1.3 Summary

Central to our proposal is the notion that words both benefit from and contribute to the creation of the message-level representation. First, words benefit from context because structure in the message-level representation facilitates the integration of elements and relations consistent with the contextually evoked frame. Further, because lexical-level expectancies are based, among other sources, on the frames active in working memory, we expect that words whose semantic contribution can easily be accommodated by the activated frame will be easier to process than words unrelated to the active frame. Second, words contribute to the message-level representation by providing cues for addressing knowledge in long term memory and for the proper construal of current conceptual content. Thus we believe that the relationship between a word and its surrounding context is multifold. This relationship involves the ways in which individual words add to the cognitive models active in working memory and how individual words can prompt the construction of new models. This view predicts that scenarios which occasion frame-shifting present a challenge to the processor which differs from that presented by lexical violations consistent with the currently active frame.

To test this hypothesis, we manipulated the relationship between sentence final words and their preceding contexts and compared word-by-word reading times for words which triggered frame-shifting, with the reading times for lexical violations (i.e., equally unexpected words) consistent with a contextually evoked frame. The primary question was whether the time needed for sentential integration of unexpected words would vary as a function of the *type* of unexpected ending. Because one-line jokes provide such an excellent source of identifiable frame-shifts, we began by contrasting participants' self-paced reading times on jokes and nonjoke controls. Further, because jokes are designed so that the listener either "gets it" or she does not, they provide a clear criterion of message-level processing. Consequently, in Experiments I and II, we compared participants' self-paced reading times for the last word of jokes with those for nonjoke controls. In Experiment III, we compared frame-shifting in jokes with frame-shifting in nonhumorous sentences that seemed to require the same sort of semantic reanalysis as the jokes.

2 Experiment I

Experiment I contrasted participants self-paced reading times for sentences which ended either as jokes, or with an unexpected word consistent with the active frame. Because joke endings presumably trigger frame-shifting, our hypothesis predicts a difference in reading times for jokes and nonjokes. In contrast, if sentential integration is driven only by factors that determine how expected a word is in a given context, reading times for jokes and nonjokes should be similar.

2.0.1 Stimulus Pretesting: Production Data

Initially, 120 one-line jokes were assembled from joke books. The stimuli were chosen so that understanding the jokes required semantic reinterpretation of meaning established earlier in the sentence. In all cases the disjunctor (the point at which the reader could, in principle,

Low Constraint Sentence My husband took the money we were saving to buy a new car and blew it all at the *Nonjoke Ending*: tables. (Cloze Probability: 3.6%) *Joke Ending*: movies. (Cloze Probability: 3.6%)
High Constraint Sentence I asked the woman at the party if she remembered me from last year and she said she never forgets a *Nonjoke Ending*: name. (Cloze Probability: 3%) *Joke Ending*: dress. (Cloze Probability: 0%)

Table 1: Sample Materials.

realize the joke) was a sentence-final noun. Jokes which fit these criteria were normed on an off-line cloze task (Bloom & Fischler, 1980). By looking at the most popular response on the cloze task we were able to establish the default nonjoke interpretations of the joke fragments. One entailment of this approach is that the operation of a general frame-based comprehension system should also be reflected in production data. This predicts a certain level of consistency in speakers' responses to fill in the blank tasks such as the cloze procedure. In fact, responses on the cloze task could be classified on the basis of consistency with evoked frames.

Moreover, stimulus pretesting revealed considerable variability in the cloze probabilities of the most popular responses (viz. the default interpretation). For instance, in sentence fragment (4) the most popular response was produced by 81% of the participants.

(4) I asked the woman at the party if she remembered me from last year and she said she never forgets a (face 81%).

However, for sentence fragment (5), participants produced a wide array of responses so that the cloze probability of even the most popular response for this sentence fragment was only 18%.

(5) My husband took the money we were saving to buy a new car and blew it all at the (casino 18%).

Consequently, sentence fragments were divided into two groups: High Constraint contexts, sentence fragments in which the cloze probability of the most popular response was greater than 40%, and Low Constraint Contexts (less than 40%). Stimuli were constructed by pairing each sentence fragment with one of two Ending Types: Joke Endings – endings which require the reader to initiate frame-shifting, and Nonjoke Endings – words which fit with the contextually evoked frame. Because all of the Joke Endings had a low cloze probability (less than 5%), we chose Nonjoke Endings similarly low in cloze probability. Mean cloze probability for joke and nonjoke endings was 2.0% (s.d.=1.8) and 3.8% (s.d.=3.5), respectively. Further, the mean length (6.5 characters, s.d.=2) and frequency (87 per million, s.d.=146) of sentence final words was not reliably different in jokes and nonjokes, nor as a function of sentence constraint. Examples of materials are listed in table 1.

To test the hypothesis that frame-shifting incurs a processing cost, we performed a word-by-word reading time study to compare the relative difficulty of reading jokes and nonjokes. By holding cloze probability constant across the two conditions, the ending type manipulation affects the extent to which participants' higher-level contextual expectancies have been violated. Our frame-based model predicts jokes should take longer to process because they require the semantic reanalysis involved in frame-shifting. In contrast, if the difficulty of sentential integration is driven by other factors that influence judgments of the conditional probability of a word in a particular context, low-cloze endings should be equally difficult to process, whether or not they are Jokes.

2.1 Methods

Participants. Sixteen UCSD undergraduates participated in fulfillment of Cognitive Science or Psychology course requirements.

Task. Participants' task was to read sentences for comprehension in a self-paced, word-by-word reading task, using the moving window paradigm (Just, Carpenter, & Wooley, 1982). In this paradigm, each sentence initally appears to be constructed of dashes. When the participant presses the space bar, the sentence's first word appears in the dashes; with each subsequent press of the space bar, the next word appears and the preceding word reverts to dashes. Consequently, participants can see the length of the sentence, and the relative lengths of the words, but they can only read one word at a time. The computer records the time between depressions of the space bar as the reading time for the word displayed during that interval.

For each sentence, a true/false comprehension question was triggered by the space bar depression which followed the appearance of the last word. The comprehension question was displayed on the screen in its entirety. Participants were told that the experiment concerned how well people dealt with various sorts of anomalies when reading, and that their task was to read sentences and try for a perfect score on the comprehension questions. Additionally, they were told to read at a comfortable pace, and not to proceed to the comprehension question until they understood the sentence.

Materials. Each participant read 30 sentences with a Joke Ending, 15 in each of the Sentence Constraint classes; and 30 sentences with a low-cloze Nonjoke Ending, 15 in each constraint class. There were two lists of stimuli varied between participants, designed so that while no individual participant read the same sentence fragment twice, when collapsed across participants, each Joke Ending was preceded by the exact same sentence fragment as its Nonjoke counterpart. The within-participants design was a 2 x 2, with two sorts of Ending Types, each preceded by sentence fragments which provided either a High or a Low Constraint context.

Besides the 60 sentences of interest, there were 100 distractor sentences, half of which contained grammatical errors and half of which served as their controls. Specifically, there were two sorts of grammatical errors, subject-verb number agreement on the verb, and pronoun case violations (on the pronouns).

Pronoun Case Errors and Controls

Sheila is looking for someone to join her/*she in a drink.

Subject-Verb Agreement Errors and Controls

Every Monday she mows/*mow the lawn.

In each list, 25 of the grammatical errors were verb agreement errors and 25 pronoun case violations; the other 50 distractor sentences were designed to serve as controls appropriate

Nonjokes	93%
Jokes	93%
Filler	93%

Table 2: Experiment I: Comprehension Scores in Percent Correct

for verb agreement and pronoun case errors, respectively.

Comprehension Questions. Participants were required to answer True/False comprehension questions after each sentence. These questions were included to encourage participants to comply with the instructions to read and comprehend the sentences. However, they were quite simple, and could be answered by paying minimal attention to the scenarios described in the sentences (see below).

Sentence Fragment

By the time Mary had had her fourteenth child, she'd finally run out of names to call her

Nonjoke Ending: pets.

Joke Ending: husband.

Comprehension Probes

Mary had run out of names for her pets. (T/F)

Mary had run out of names for her husband. (T/F)

2.2 Results

2.2.1 Comprehension Scores

During post-experiment debriefing, participants' comments suggested they realized that jokes had been interspersed amongst the stimuli, and, further, that they had no trouble understanding the materials. Participants' scores on comprehension questions were high, and indicate they did indeed appreciate the meaning of the sentences (see table 2). Repeated measures ANOVA on comprehension scores for questions following Joke, Nonjoke, and Filler Materials suggested performance was equally good for all stimulus categories $MaterialsF(2,30) < 1$.

2.2.2 Reading Times

Because all of the jokes turned on the last word of the sentence, the dependent variable of interest was the reading time for the last word of experimental sentences. Reading times for each participant were trimmed at plus or minus 2 standard deviations from the mean in each cell. Less than 6 percent of all reading times were trimmed as a result of this procedure. Reading times, listed in table 3, were analyzed with repeated measures ANOVA with factors Ending Type (Nonjoke/Joke) and Sentence Constraint (Lower/Higher) and Participants as the random factor (referred to below as F1). In the items analysis (F2), Sentence Constraint (Lower/Higher) was treated as a between-items factor, while Ending Type was treated as a within-items factor.

As predicted, participants spent an average of 132 msec longer on the last word of joke stimuli than on their nonjoke counterparts [Ending Type F1(1,15)=71.4, $p < .05$;

	Low Constraint Mean (s.d.)	High Constraint Mean (s.d.)	Mean (s.d.)
Nonjoke	795 (233) ms	640 (162) ms	718 (388) ms
Joke	856 (265) ms	843 (320) ms	850 (523) ms
	825 (249) ms	741 (272) ms	

Table 3: Experiment I: Reading Times for the Last Word of Jokes and Nonjokes

F2(1,58)=10.23, $p < 0.01$]. However, the difference between Joke and Nonjoke reading times was far more pronounced for the High Constraint sentences than the Low Constraint sentences. The interaction between Sentence Constraint and Ending Type was reliable in the Participants analysis, but marginal in the Items analysis [F1(1,15)=6.32, $p < 0.05$; F2(1,58)=3, $p = 0.09$]. Post-hoc t-tests of Ending Type, revealed a reliable joke effect in High Constraint sentences [F1(1,15)=9.65, $p < 0.01$; F2(1,58)=9.57, $p < 0.01$], but not in Low Constraint sentences [$F1(1,15) = 1.66, p = 0.22; F2(1,58) < 1$].

2.3 Discussion

Although none of the sentence final words in these materials was predictable from the preceding sentence context, people spent appreciably longer on the last word of sentences which ended as Jokes than those which ended as Nonjokes. Both Joke and Nonjoke endings had an average cloze probability of less than 5%. However, while Nonjoke endings were constructed so as to be congruent with the frame evoked by context, Joke endings were thought to require the initiation of frame-shifting. The main effect of Ending Type on reading times for sentence final words is thus consistent with the hypothesis that semantic reanalysis necessary to integrate the Joke Ending into the message-level representation exerts a processing cost which exceeds that associated with integrating a word that is equally unexpected but congruent with the contextually evoked frame.

However, the frame-shifting effect was not ubiquitous, but rather confined to the High Constraint stimuli, the sentences that suggested a particular lexical item as the best completion for the sentence. In spite of the fact that participants in Experiment I never encountered the *expected* completion, results suggest that High Constraint sentence contexts aided sentential integration of unexpected items consistent with the message-level frame, while making frame-shifting more difficult. Because High Constraint sentences were those sentences which strongly suggested a particular lexical item on the production task, it may be that the production task is an indirect index of the likelihood a given reader will commit to a particular frame.

Thus while High Constraint sentences allow the reader to commit to a frame with which to structure language input, Low Constraint sentences engender less frame commitment. In any case, Low Constraint sentence fragments seem to elicit a message-level representation that makes it equally difficult to integrate Joke and Nonjoke endings. One possible reason for the absence of an Ending Type effect in Low Constraint stimuli is that the contextually evoked frame is highly schematic and has very few constraints on its slots. Consequently, frame-shifting might be required to some extent for the integration of Joke and Nonjoke endings alike.

Another possibility is that the participants and the experimenters were more likely to construct the same nonjoke frame for High Constraint stimuli than they were for Low Constraint stimuli. That is, because High Constraint sentence fragments suggest a particular completion, participants in Experiment I were likely to utilize the frame that was consistent with the nonjoke ending chosen for those sentences. Further, because people's interpretation of Low Constraint sentences is more variable (by definition), it is possible the nonjoke endings provided by the experimenters did not match the nonjoke frame constructed by participants. If this were the case, frame-shifting would be required in both the joke and the nonjoke sentences. In any case, Low Constraint sentence fragments seem to elicit a message-level representation that makes it equally difficult to integrate the Joke and the Nonjoke Endings.

3 Experiment II

To ensure that participants were comprehending the experimental stimuli, we increased the difficulty of the comprehension questions for the Joke and the Nonjoke stimuli. In the original study, comprehension questions either paraphrased the stimulus sentence ("True" questions), or conveyed a slightly different meaning than the stimulus sentence ("False" questions). In Experiment II, we used comprehension questions that required the reader to have drawn a key inference about the scenario in the stimulus sentence. It was hoped that this would provide a more stringent test of comprehension, as well as encouraging participants to read sentences more closely.

3.1 Methods

Participants. 32 UCSD undergraduates participated for Cognitive Science or Psychology course credit. All participants were native speakers of English, although several were fluent in other languages as well.

Task. Participants' task was to read sentences in a self-paced word-by-word reading paradigm as used by Just, Carpenter, & Wooley (1982). As in the previous experiment, participants were instructed to read sentences for comprehension. Immediately after each sentence, participants were required to answer a true/false question about the sentence they had just read. While all participants were encouraged to strive for a perfect score on comprehension questions, participants in the Difficult Question condition were told that some questions concerned the implications of the preceding statement and might require making judgments about what the speaker was "getting at."

Materials. Materials in Experiment 2 were identical to those used in Experiment 1 (described above), with the exception of the comprehension questions following Joke and Nonjoke stimuli. Half of the participants answered the "Simple" questions used in Experiment 1 (essentially replicating that study), while the other half of the participants answered "Difficult" questions. As described above, difficult questions were designed to ensure that the reader had drawn a key inference which enabled her to "get" the joke. For example, after (6), half of the participants were given Simple Questions (like (7)) from Experiment

	Simple	Difficult
Nonjoke	94 %	89 %
Joke	97 %	71 %
Fillers	94 %	94 %

Table 4: Experiment II: Comprehension Scores in Percent Correct

I which tested whether they were attending to the stimuli, and placed a special emphasis on whether participants had seen sentence final words. The other half of the participants were given Difficult Questions (like (8)) which tested whether they had made a key inference necessary to get the joke.

(6) I asked the bartender for something cold and full of rum, and he recommended his wife.

Simple Question

(7) The bartender recommended his wife. (T/F)

Difficult Question

(8) The bartender's wife is a frigid lush. (T/F)

3.2 Results

3.2.1 Comprehension Scores

Comprehension scores were again quite high, both for the participants who answered Simple Questions, and for those who answered the more Difficult Questions (see table 4). A repeated measures ANOVA on comprehension scores with between-participants factor Difficulty (Simple/Difficult) and within-participants factor Materials (Nonjoke/Joke/Fillers) yielded reliable main effects of Difficulty, Materials, and an interaction between the two [Difficulty F(1,30)=84.19, $p < 0.001$; Materials F(1,30)=35.08, $p < 0.001$; Difficulty x Materials F(1,30)=61.94, $p < 0.001$].

Participants who were given Simple comprehension questions scored significantly higher than did participants given the more difficult questions. Moreover, the interaction between Difficulty and Materials was due to the fact that comprehension performance on the Simple questions was equally good for Nonjoke and Joke stimuli, while on the Difficult questions participants' performance was reliably better for questions concerning the Nonjoke as opposed to Joke stimuli.

3.2.2 Reading Times

Reading times were trimmed at plus or minus 2 standard deviations from the mean, thereby affecting 6 percent of the total reading times. Reading Times were then analyzed using repeated measures ANVOA with Groups (Simple Questions/Difficult Questions) as a between-participants variable, and both Sentence Constraint (Lower/Higher) and Ending Type (Nonjoke/Joke) as within-participant variables. In the items analysis, Sentence Constraint was used as a between-items variable, and Groups and Ending Type were used as within-items

Group A: Simple Questions	Low Constraint Mean (s.d.)	High Constraint Mean (s.d.)	Mean (s.d.)
Nonjoke	718 (265) ms	634 (261) ms	676 (262) ms
Joke	783 (356) ms	782 (301) ms	782 (324) ms
	750 (311) ms	708 (287) ms	

Table 5: Experiment II: Reading Times for Group A

Group B: Difficult Questions	Low Constraint Mean (s.d.)	High Constraint Mean (s.d.)	Mean (s.d.)
Nonjoke	740 (294) ms	643 (349) ms	691 (321) ms
Joke	786 (313) ms	836 (422) ms	811 (366) ms
	763 (300) ms	739 (393) ms	

Table 6: Experiment II: Reading Times for Group B

variables. Reading times for the group who answered Simple Questions are listed in table 5, while reading times for the group who answered Difficult Questions are listed in table 6.

The results of both groups in Experiment II largely replicate those of Experiment I. While there was a trend for longer reading times in the group who answered the more difficult questions, the overall effect of Question Group was null [$F1[participants](1,30) < 1$; $F2[items](1,116) < 1$]. As in the first experiment, there was an overall effect of Ending Type with participants spending an average of 113 ms longer on the Jokes [F1(1,30)=13.69, $p < 0.001$; F2(1,116)=15.08, $p < 0.001$. Moreover, we again observed an interaction between Ending Type and Sentence Constraint due to a larger, more robust Joke effect in the Higher Constraint sentences than in the Lower Constraint sentences [F1(1,30)=5.34, $p < 0.05$; F2(1,116)=3.95, $p < 0.05$]. Collapsing across Question Groups, post-hoc t-tests reveal a reliable effect of Ending Type among the High Constraint stimuli [F1(1,31)=18.35, $p < 0.001$; F2(1,29)=13.32, $p < 0.001$], with people spending an average of 171 milliseconds longer reading the Joke endings than the Nonjoke endings. We observed a similar trend among Low Constraint stimuli, with people spending an average of 55 milliseconds longer reading the Joke Endings. However this effect was not significant [F1(1,31)=2.11, p=.15; F2(1,29)=1.56, p=.22].

3.3 Discussion

Comprehension scores in Experiment II suggest that while participants' scores in Experiment I may have overestimated their comprehension level, participants did indeed understand the bulk of what they were reading. As in Experiment I, participants in the Simple Question Group scored 90% or above in all categories (Jokes, Nonjokes, and Fillers). In contrast, participants in the Difficult Question Group scored slightly lower on questions following nonjokes (89% correct) and on jokes (71% correct). Because both groups scored equally well on questions following filler materials – questions which were identical for both groups – lower scores probably reflect increased difficulty of the questions rather than group differences in reading ability. Further, lower scores observed in the Difficult Group for questions following jokes suggests that these stimuli place some additional demands on the comprehension system relative to nonjokes, designed to be congruent with frame-level expectations.

Furthermore, reading times observed in Experiment II showed the same patterns as those in Experiment I. Once again, reading times for joke endings were longer than for nonjoke controls. Moreover, we observed the same interaction between sentence constraint and ending type. As in Experiment I, the ending type effect was larger and more robust in high constraint stimuli than in low. Although the ending type effect was not significant in low constraint stimuli, the trend towards longer reading times for joke endings was observed in both the Simple Question Group and the Difficult Question Group. These data suggest that an effect of frame-shifting may have been present in *some* of the low constraint stimuli, but not all of them.

Further, while reading times for joke endings in high and low constraint sentences were quite similar, reading times for nonjoke endings in high constraint sentences were always shorter than for nonjoke endings in low constraint sentences. This pattern is what one would expect if the nonjoke endings for low constraint sentences were sometimes consistent with frame-level expectations, and sometimes not. If this is indeed the case, then longer reading times for low constraint nonjoke endings might reflect frame-shifting in some *nonjoke* stimuli rather than an absence of frame-shifting in low constraint jokes.

4 Experiment III

4.0.1 Experiment IIIA

Experiments I and II suggested that, at least in High Constraint sentence contexts, words which trigger frame-shifting require increased processing time relative to unexpected words consistent with a contextually evoked frame. However, because the frame-shifting manipulation contrasted jokes with nonjokes, it was confounded by the fact that the joke stimuli were funnier than their controls. Consequently, Experiment IIIA was conducted to test the generality of frame-shifting, by testing whether participants' reading times for nonhumorous examples that require frame-shifting would display the same effects observed in reading times for jokes. We did so by constructing 'non-humorous' sentences that seemed to trigger frame-shifting in much the same way as did the jokes. We call sentences like (9) – that prompt the reader to shift frames on the sentence-final words – Shift stimuli.

Shift

(9) The veterans were suing the government because they had been exposed to dangerous ideas.

Using a large database of sentences that had been normed on the cloze task, we were able to find 60 sentence fragments which could be completed both with a word which triggered a shift, and with an unexpected word consistent with frame-level expectations. For example, (9) can be completed either with "ideas" (the Shift Ending), or with "toxins" (the Nonshift Ending), words which matched the Shift Endings in cloze probability, yet were consistent with the overall frame evoked by the context. Further, as in the Joke/Nonjoke materials, the 60 sentences chosen for Shift/Nonshift materials were divided into High and Low Constraint sentences based on the cloze probability of their best completions. Like their Joke/Nonjoke counterparts, there were 30 Shift/Nonshift sentences whose best completion had a cloze probability of greater than 40%, and 30 sentences whose best completion had a cloze probability of less than 40%. Cloze probability, length, and frequency of Shift and Nonshift Endings were matched, and did not differ reliably from those of the Joke/Nonjoke Endings.

In order to test the hypothesis that frame-shifting is not a unique feature of joke materials, we recorded participants self-paced reading times as they read Joke/Nonjoke and Shift/Nonshift sentences and answered questions about the materials. Although the results of Experiment II suggested that the pattern of effects was not modulated by Question Difficulty, this factor was once again varied between participants. If frame-shifting is a general process involved in comprehension, we would expect to observe a similar effect of frame-shifting in reading times for Shift/Nonshift and Joke/Nonjoke materials.

4.0.2 Experiment IIIB: List Effects for Joke/Nonjoke Stimuli

Because pilot testing suggested that the inclusion of semantic anomalies in the stimulus set could change the pattern of results obtained for the reading times for the Joke and the Nonjoke Endings, Experiment IIIB was conducted to systematically examine whether list composition used in Experiment IIIA would affect reading times for Joke versus Nonjoke Endings. Consequently, concurrent with data collection for Experiment IIIA, we collected reading time data for the same Joke/Nonjoke stimuli from a different group of 16 participants who read experimental stimuli embedded in a different list. The resultant design included two between-participant variables (Question Difficulty and Filler Content) and two within-participant variables (Ending Type and Sentence Constraint).

4.1 Methods

Participants. Sixteen UCSD undergraduates participated in ExperimentIIIA for Cognitive Science or Psychology course credit. Experiment IIIB involved the participation of an additional sixteen people from the same pool. All participants were native English speakers, although some were fluent in other languages as well. None had participated in the previous studies.

Procedure. As in experiments described above, participants' task was to read sentences in a self-paced moving window reading paradigm as used by Just, Carpenter, & Wooley (1982). Immediately after each sentence a true/false question appeared on the screen for the participant to answer. While all participants were encouraged to strive for a perfect score on comprehension questions, participants in the Difficult Question condition were told that some questions concerned the implications of the preceding statement, and might require making judgments about what the speaker might have been "getting at."

4.1.1 Experiment IIIA

Materials. Each participant read 60 sentences from the Joke/Nonjoke materials employed in Experiments I and II, as well as 60 sentences from the nonhumoruous Shift/Nonshift materials. As in the earlier studies, each participant read 30 Jokes, 15 in each of the constraint classes, and 30 Nonjokes, 15 in each of the constraint classes. Additionally, each participant read 30 Shift sentences, 15 in each of the constraint classes, and 30 Nonshift sentences, 15 in each of the constraint classes. 40 additional sentences were included as filler items, 20 Related Anomalies as in (10), and 20 Unrelated Anomalies as in (11). Related Anomalies were sentences which ended with an incongruous item semantically related to the expected item, while Unrelated Anomalies were merely sentences which ended incongruously.

Related Anomalies

(10) They asked Dave to play tennis, but he had to restring his court.

Unrelated Anomalies

(11) The businessman took the receipts to the first national accent.

4.1.2 Experiment IIIB

The intent of Experiment IIIB was to ensure that neither the inclusion of Shift and Nonshift sentences, nor of Related and Unrelated Anomalies would fundamentally alter the pattern of reading times for the Joke/Nonjoke stimuli in Experiment IIIA. Consequently, all 32 participants in Experiment IIIB read 30 Joke, and 30 Nonjoke stimuli. However, half of the participants (those in Experiment IIIA) read Joke/Nonjoke materials in the "Anomalous Filler List" described above (containing Shift and Nonshift Materials, as well as Related and Unrelated Anomalies); the other half read the 30 Joke and 30 Nonjoke stimuli embedded in the "Grammatical Filler List," in which distractor sentences consisted of 50 pronoun case violations and 50 grammatical controls.

The difficulty level of the comprehension questions for the distractor sentences was held constant across lists. However, the difficulty of the questions following the experimental stimuli (Jokes and Nonjokes) was varied as in Experiment II, with half of the participants (8 in each List Composition category) answering the original "Simple" comprehension questions, and half of the participants (8 in the Grammatical Filler List, and 8 in the Semantically Anomalous Filler List) answering the "Difficult" questions as described above.

4.2 Results and Discussion

4.2.1 Experiment IIIA: Comprehension Scores

Comprehension scores for the 16 participants who read both Joke/Nonjoke and Shift/Nonshift stimuli are listed in table 7. These scores were analyzed with repeated measures ANOVA treating participants as the random factor. Question Group was treated as a between-participants factor, and Humor and Ending Type were treated as within-participants factors. Note that Question Group refers *only* to the questions which followed the jokes: questions which followed Shift and Nonshift materials were the same for both groups. Analysis revealed a main effect of Question Difficulty [$F(1,14)=14.02, p < 0.01$] reflecting higher scores in the group that answered the Simple Questions; a main effect of Humor [$F(1,14)=18.99$;

Simple Questions				Overall Mean
Nonjoke	84%	Nonshift	93%	88%
Joke	96%	Shift	89%	93%
Humorous	90%	Nonhumorous	91%	90%
Difficult Questions				Overall Mean
Nonjoke	88%	Nonshift	88%	88%
Joke	65%	Shift	77%	77%
Humorous	76%	Nonhumorous	82%	82%

Table 7: Comprehension for Nonjoke, Joke, Shift, and Nonshift Materials

Humorous Materials	Mean (s.d.) sem	**Nonhumorous Materials**	Mean (s.d.) sem	Overall Mean (s.d.) sem
Nonjoke	795 (280) 50	Nonshift	684 (238) 42	740 (264) 33
Joke	902 (384) 68	Shift	785 (354) 62	844 (372) 46
Mean	849 (338) 42	Mean	735 (303) 38	

Table 8: Reading Times (in ms) for Sentence Final Words in Humorous and Nonhumorous Materials

$p < 0.001$], reflecting participants' lower scores on questions about Joke/Nonjoke materials; and a main effect of Ending Type [F(1,14)=5.8; $p < 0.05$], reflecting lower scores for questions following the Nonjoke/Nonshift materials than Joke/Shift materials which required frame-shifting.

However, comprehension scores were also characterized by interactions between Question Group and Humor [F(1,14)=13.69,$p < 0.01$], between Question Group and Ending Type [F(1,14)=27.57, $p < 0.0001$], and between Question Group, Humor, and Ending Type [F(1,14)=39.25, $p < 0.0001$]. Scores listed in table 7 suggest the root of these interactions is the poor performance by participants in the Difficult Question Group on Jokes, as well as the low scores on Nonjokes in the Simple Question Group.

4.2.2 Experiment IIIA: Reading Times

Reading times were analyzed with a repeated measures ANOVA with Question Groups as a between-participants factor, and Humor (Nonjoke-Joke/Nonshift-Shift), Sentence Constraint (Lower/Higher), and Ending Type (Nonshifting/Shifting) as within-participants factors. We observed main effects of Humor [F(1,14)=27.26, $p < 0.001$], with participants spending 114 ms longer on Nonjoke-Joke materials, and Ending Type [F(1,14)=10.32, $p < 0.01$], with participants spending approximately 100 ms more on words which triggered frame-shifting. However, there was no interaction [$HumorxEndingTypeF(1,14) < 1$]. Reading Times for the last word of Joke, Nonjoke, Shift, and Nonshift Materials can be found in table 8.

Further, while the overall effect of Ending Type was similar in the humorous and nonhumorous materials, we did observe an interaction between Humor and Sentence Constraint

Nonjoke/Joke Materials		**Nonshift/Shift Materials**	
	Mean (s.d.)		Mean (s.d.)
Low	874 (314)	Low	714 (294)
High	824 (364)	High	756 (316)

Table 9: Reading Times (in ms) for Sentence Final Words of Low and High Constraint Contexts in Humorous and Nonhumorous Materials

[F(1,14)=8.57,$p < 0.05$]. This interaction reflects the fact that the effect of Sentence Constraint on reading times for the Joke/Nonjoke materials was quite similar to that observed in Experiments I and II, with slightly longer reading times for the last words of Low Constraint sentences. In contrast, the effect of Sentence Constraint on reading times for the Shift/Nonshift materials was in the opposite direction – with longer overall reading times for the last word of High Constraint sentences. Mean reading times for the last word of High and Low Constraint sentences can be found in table 9.

There was no overall effect of Question Group in the subjects analysis [$F1(1, 14) < 1$]. Moreover, while Question Group did not reliably interact with any of the other variables, there was a trend for a three-way interaction between Question Group, Constraint, and Ending [F1(1,14)=3.52, p=0.08]. In the group that answered Simple Questions, we observed a pattern of results not unlike that in the earlier studies: a trend for longer reading times on words which triggered frame-shifting in Low Constraint sentences, and a larger, more robust, difference among High Constraint sentences. However, in the group that answered Difficult Questions, there was a large, robust effect of frame-shifting in Low Constraint sentences, and a trend to spend 40 ms longer on words which triggered frame-shifting in High Constraint sentences. Moreover, this three-way interaction was further qualified by a trend for a four-way interaction between Question Group, Humor, Sentence Constraint, and Ending Type [F(1,14)=3.10, p=0.10].

The four-way interaction results from the behavior of the Difficult Question group on the Joke/Nonjoke materials. In the group who answered Simple Questions, we observed an interaction between Sentence Constraint and Ending Type which was quite similar to that the previous two studies. In this group, both for the Joke/Nonjoke and the Shift/Nonshift materials, the effect of frame-shifting derives almost exclusively from the High Constraint materials. In contrast, in the group that answered Difficult Questions, the Ending Type effect on Shift/Nonshift materials was approximately the same size in High and Low constraint contexts. Moreover, unlike any of our previous results, in High Constraint sentences we observed longer reading times for *Nonjoke* than Joke Endings. As noted above, this group of participants spent longer reading the last word of Jokes than Nonjokes in the *Low* Constraint sentences.

4.2.3 Experiment IIIB: Comprehension Scores

As table 10 suggests, participants seemed to comprehend the materials. However, scores on the Difficult Questions in the Anomaly Fillers Group were quite low for the Jokes (only 65% correct). The presence of so many semantic anomalies in the stimulus set may have discour-

Grammatical Fillers Group	Simple Questions	Difficult Questions
Nonjokes	91%	91%
Jokes	93%	76%
Fillers	92%	90%
Anomaly Fillers	Simple	Difficult
Group	Questions	Questions
Nonjokes	83%	88%
Jokes	96%	65%
Fillers	91%	91%

Table 10: Experiment III: Comprehension Scores in Percent Correct

aged participants from fully processing the meaning of the Jokes in these lists. Repeated measures ANOVA with between-participants factors Difficulty (Simple/Difficult) and Fillers (Grammatical/Anomaly), and within-participants factor Materials (Nonjokes/Jokes/Fillers) suggested a reliable effect of Difficulty, due to better performance by participants in the Simple question condition 91% versus 84% [Difficulty F(1,28)=20.64, $p < 0.001$]. There was a reliable effect of Filler Set, due to better performance by participants in the Grammatical than the Anomaly Filler Set (90% versus 86% correct) [Fillers F(1,28)=5.82, $p < 0.05$]. Further, participants performed best on comprehension probes which followed Fillers (92%), next on probes which followed Nonjokes (88%), and worst on probes which followed Jokes (83%) [Materials F(2,56)=29.89, $p < 0.001$].

Reliable interactions between Materials and Difficulty [F(2,56)=66.95, $p < 0.001$], and between Materials, Difficulty and Fillers [F(2,56)=6.99, $p < 0.01$], suggest that the main effect of Materials is driven by poor performance on Difficult questions which followed Jokes, especially when embedded in the Anomaly Filler set (as discussed in Experiment IIIA). Performance on questions following Nonjokes and Filler items was not modulated by question Difficulty; however, performance on probes following Jokes dropped from 95% in Simple questions to 71% on Difficult questions. Further, the three-way interaction suggests that participants' comprehension of the Jokes was particularly compromised by embedding them in the Anomaly filler set. Participants in Grammatical Fillers condition correctly answered an average of 71% of Difficult questions which followed Jokes. However, participants in the Anomaly fillers condition averaged a score of only 65% on the same questions.

4.2.4 Experiment IIIB: Reading Times

Although there was no overall effect of Question Group in the subjects analysis [F1(1,28)=1.5, p=.23], Question Group *was* significant as a within-items variable in the items analysis [F2(1,58)=28.82, $p < 0.0001$]. Marginal means listed in table 11 indicate that people spent longer reading Joke and Nonjoke Endings when they had to answer difficult questions. Moreover, we also observed a Question Group x Filler Type interaction [F1(1,28)=4.4, $p < 0.05$; F2(1,58)=64.4, $p < 0.00001$] due to the fact that increasing question difficulty led to *longer* overall reading times among groups who read experimental stimuli along with the grammatical filler set, while it led to *shorter* overall reading times among groups who read experimental

	Grammatical Fillers Group Mean (s.d.)	Anomaly Fillers Group Mean (s.d.)	Mean (s.d.)
Simple Questions Group	722 (172) ms	901 (331) ms	811 (277) ms
Difficult Questions Group	1117 (572) ms	797 (342) ms	957 (495) ms
Mean (s.d.)	920 (464) ms	850 (338) ms	

Table 11: Experiment III: Overall Reading Times for Joke and Nonjoke Stimuli in the Four Sub-Groups

Grammatical Fillers Groups	Simple Questions Group	Difficult Questions Group	**Anomaly Fillers Groups**	Simple Questions Group	Difficult Questions Group
	Mean (s.d.)	Mean (s.d.)		Mean (s.d.)	Mean (s.d.)
Low	727 (208) ms	1046 (560) ms	Low	895 (276) ms	853 (355) ms
High	716 (133) ms	1187 (594) ms	High	907 (387) ms	740 (330) ms

Table 12: Sentence Constraint Effect in Four Sub-Groups: Reading Times for Sentence Final Words in Low and High Constraint Sentences

stimuli along with the anomaly filler set (see table 11).

Note that longer overall reading times in the grammatical filler set also corresponded to the groups who did well on the comprehension probes, while the shorter overall reading times in the Anomalous Filler Group can be attributed to the people with the lowest comprehension scores. These data suggest embedding the joke/nonjoke stimuli in semantic anomalies discouraged some participants from fully processing them. Further, the fact the attenuated joke effect was associated both with decreased comprehension scores, and with shorter overall reading times indicates the cognitive nature of the original effect.

Further, while there was no overall effect of Sentence Constraint [$F1(1,28) < 1; F2(1,58) < 1$], there was a 3-way interaction between Question Group, Filler Group, and Constraint [F1(1,28)=5.51, $p < 0.05$; F2(1,58)=4.8, $p < 0.05$]. The group that read experimental stimuli amongst grammatical fillers and answered Simple Questions had shorter reading times for the last word of High Constraint Sentences than for Low Constraint Sentences, a pattern shared by the group who answered Difficult Questions and read experimental stimuli amongst the Anomaly Filler set. In contrast, the group who answered Simple Questions and read stimuli embedded in the Anomaly Filler set showed shorter reading times for the last word of Low Constraint sentences; as did the group who answered Difficult Questions and read experimental stimuli amongst the Grammatical Filler set (see table 12).

As in Experiments I and II, there was an overall effect of Ending Type [F1(1,28)=23.99, $p < 0.0001$; F2(1,58)=29.52, $p < 0.0001$], with people spending an average of 181 milliseconds longer reading the Joke Endings than their Nonjoke counterparts. Moreover, the Ending Type effect in Experiment IIIb was twice as large among groups who read experimental stimuli embedded in the Grammatical Filler set as it was in groups (those discussed in

	Grammatical Filler Group	Anomaly Filler Group	Overall
	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)
Nonjoke	785 (338) ms	795 (280) ms	791 (308) ms
Joke	1053 (535) ms	902 (384) ms	978 (468) ms

Table 13: Experiment III: Ending Type x Filler Group

Grammatical Fillers

Simple Question Group	Low Constraint Mean (s.d.)	High Constraint Mean (s.d.)	Difficult Question Group	Low Constraint Mean (s.d.)	High Constraint Mean (s.d.)
Nonjokes	635 (85) ms	642 (76) ms		882 (377) ms	980 (499) ms
Jokes	819 (258) ms	790 (140) ms		1210 (684) ms	1394 (641) ms

Anomaly Fillers

Simple Question Group	Low Constraint Mean (s.d.)	High Constraint Mean (s.d.)	Difficult Question Group	Low Constraint Mean (s.d.)	High Constraint Mean (s.d.)
Nonjokes	889 (287) ms	763 (217) ms		766 (170) ms	764 (420) ms
Jokes	901 (286) ms	1051 (476) ms		941 (472) ms	716 (237) ms

Table 14: Experiment III: 4-way Interaction

Experiment IIIa) who read the same stimuli embedded in the Anomaly Filler set (see table 13) (Filler Group x Ending Type [F1(1,28)=4.43, $p < 0.05$; F2(1,58)=5.91, $p < 0.05$]. Further, the data in table 14 suggest the Ending Type effect may be further qualified by the presence of a four-way interaction between Question Group, Filler Set, Sentence Constraint and Ending Type suggested by the subjects analysis [F1(1,28)=3.98, p=0.055].

This trend towards a four-way interaction reflects the deviant pattern of reading times displayed by the group that read experimental stimuli embedded in the anomaly filler set, and who answered difficult questions. Unlike the members of the other groups (and, incidentally, participants in Experiments I and II) among whom the Joke effect was most robust among High Constraint sentences, participants in the Difficult Question/ Anomaly Fillers spent the *least* time reading High Constraint Joke Endings. Perhaps not coincidentally, this group of participants also performed the worst on the comprehension questions which followed the jokes, averaging only 65% correct on the True/False judgment task.

5 General Discussion

5.1 Ending Type

In Experiments I and II we tested the psychological reality of frame- shifting by comparing self-paced reading times in sentences which ended either as Jokes or as Nonjokes. Further, in Experiment III we tested whether frame-shifting could be detected in 'nonhumorous' stimuli. Although both the Joke and the Nonjoke Endings violated the sorts of conscious expectancies that influence people's behavior on production measures such as the cloze task, they each did so in a different way. Nonjoke Endings were designed to be consistent with the frames evoked by the preceding sentence fragment, but not with the reader's specific lexical expectations. Joke Endings on the other hand, were designed to violate both the lexical- and frame- level expectations of the reader. Similarly, Nonshift Endings were unexpected words consistent with the frames evoked by the preceding sentence fragment, while Shift Endings were designed to trigger frame-shifting.

The psychological reality of frame-shifting is supported by the presence of an Ending Type effect in which people spent longer reading Joke than Nonjoke Endings, in all three experiments reported above. These data are consistent with the hypothesis that words that prompt the language user to initiate frame-shifting exert a processing cost which exceeds that incurred by unexpected words that are nonetheless consistent with message-level expectations. Moreover, the results of Experiment III suggest that frame-shifting is not confined to one-line jokes, but is a more general phenomenon. Nonhumorous examples which required frame-shifting also elicited increased reading times relative to equally unexpected controls designed to be consistent with contextual expectations at the message level.

Participants' sensitivity to the Ending Type manipulation thus suggests that the frame-level expectations that determine behavior on the off-line production task, also influenced processes invoked in our on-line comprehension task. In fact, the only subcondition in which we failed to observe longer reading times for Jokes was in the condition in which participants' performance on the comprehension probes was only 65%. Taken together, these findings suggest that increased reading times for joke endings reflect cognitive activity necessary to get the jokes.

5.2 Sentence Constraint

Another robust finding of these experiments was the importance of Sentence Constraint. While the pattern of results was modulated somewhat by global factors such as Question Difficulty and List Composition, in general the Joke effect was only evident in our High Constraint sentences. Presumably High Constraint sentences *are* high constraint because they allow the reader to commit to a particular frame to structure the message-level interpretation, a notion we refer to (after Lange & Dyer, 1989) as *frame commitment*. Because Low Constraint sentences engender less frame commitment, there is less of a difference between the demands of integrating Joke and Nonjoke endings. Consequently, reading times for the last word of low constraint sentences were generally longer than for high constraint sentences.

The nature of the observed interaction between ending type and sentence constraint might appear to conflict with a result reported by Schwanenflugel & Shoben (1985). These investigators had found that recognition of expected but not unexpected endings was facilitated by high constraint sentences, while low constraint sentences facilitated both. However, while Schwanenflugel & Shoben's Unexpected condition was quite similar to our Nonjoke and Nonshift conditions, they had no counterparts for our Joke and Shift conditions. Further, while Schwanenflugel & Shoben compared word processing in biased versus neutral contexts, we compared word processing which requires revision of the message-level representation versus processing which does not. These differences reflect their inherent interest in context *effects* as opposed to the construction of the contextual representation itself.

Our findings also have a parallel in the sentence processing literature in which eye tracking data suggest 'commitment' to a particular parse influences how willing readers are to abandon their initial misanalysis (Pickering & Traxler, 1995). Comparison of people's gaze durations of (12) and (13) to sentences which included a disambiguating comma suggested that, initially, participants misanalyzed "the magazine" as being the the object of the verb "edited" in (12), and of the verb "sailed" in (13) (Pickering & Traxler, 1995).

(12) As the woman edited the magazine about fishing amused all the reporters.

(13) As the woman sailed the magazine about fishing amused all the reporters.

Further, their results also suggested that people were more willing to abandon this initial misanalysis in (13) than in (12), presumably because the misanalysis in (12) is more plausible. Comparison of gaze duration for individual words in (12) and its unambiguous control suggested readers began to experience processing difficulty at the verb "amused." However, a similar comparison in (13) suggested readers began to experience processing difficulty at "magazine." The finding that semantic plausibility affects readers' commitments to syntactic misanalyses suggests that the garden pathing observed by Pickering & Traxler was semantic as well as syntactic.

Pickering & Traxler report similar results for sentences in which the plausibility manipulation involved the use of contextually motivated nonstandard meanings for words. For example, they contrasted gaze durations in sentences like (14) alternatively preceded by a context which mentioned bronze statues *of* particular professors, or by a context in which the janitor was polishing statues *for* particular professors.

(14) While the janitor was polishing the professor of physical chemistry prepared a lecture in his office.

Again, the pattern of gaze durations suggested readers were less willing to abandon the polish-the-professor garden path when context suggested a plausible reading for this misanalysis.

If the mechanism underlying Pickering & Traxler's commitment effects has a functional role that goes beyond the scope of the parser, we might expect to see similar effects in misanalyses of syntactically *unambiguous* sentences such as the jokes in the experiments described above. Because jokes are syntactically unambiguous sentences that have been misanalyzed on semantic and/or pragmatic grounds, the greater processing cost of frame-shifting in High Constraint Jokes could have been due to this type of commitment effect. Perhaps readers were more hesitant to abandon their initial frames in High than Low Constraint sentences. Further, the similarity between data reported by Pickering & Traxler (1995) and the semantic garden pathing in jokes highlight the influence of message-level processing on measures traditionally used to assess sentence processing.

While psycholinguists have traditionally approached meaning construction from the perspective of how low-level processing of words is influenced by the develop*ing* message-level representation, these data suggest we also need to consider how the lexical processing influences the develop*ment* of the message-level representation. Effects that have been attributed to mechanisms (such as spreading activation in the lexicon) that utilize contextual information to facilitate or inhibit low-level processes such as word recognition and lexical access, may instead (or also) reflect differences in the complexity of high-level processes of meaning construction. We suggest that whatever else it may entail (e.g. the activation of representations in the lexicon), lexical processing triggers the creative construction of cognitive models in working memory. Consequently, contextual manipulations modulate the complexity of the processes initiated by a given word, as well as the complexity of any obligatory processes in a word's processing.

In the next section we examine in more detail the frame-shifting needed to understand some of jokes used in Experiments I-III. We point to some of the difficulties of a frame-based comprehension system, showing how our data highlight the dynamically adaptive nature of the interpretive process that far exceeds that envisioned (and occasionally implemented) in the pioneering proposals of frames (e.g. Minsky, 1975; Schank & Abelson, 1977; and Rumelhart, 1980). Finally, we motivate a more flexible system which incorporates ideas from cognitive semantics to ameliorate some of the limitations of traditional frame-based comprehension.

6 Towards a unified model of saying and meaning

6.1 Frame-Based Models of Comprehension

The constructivist model of comprehension sketched in the introduction relies extensively on the idea that the meaning construction process recruits frames. Further, the semantic reorganization involved in frame-shifting would seem to require that the message-level representation have at least some properties of frames. In particular, the computational challenge of connecting an initial interpretation to the reinterpretation seems to require the representation of causal and relational information, slot/filler organization, and the existence of default values.

However, traditional implementations of frames have fallen out of favor for being overly brittle. Wilensky (1986), for example, argues that scripts are rigid data structures that cannot accommodate events that are out of the ordinary. While the restaurant script works well when the events in the restaurant coincide with those in the script, there are an infinite number of unexpected restaurant events which it is incapable of representing. Similarly, in many of the jokes we used in Experiments I-III, knowledge of typical scenarios is *necessary* for comprehension, but far from sufficient.

Example (15), for instance, suggests the import of default assumptions by violating one.

(15) Everyone had so much fun diving from the tree into the swimming pool we decided to put in a little water.

The word "water" here is surprising, not because it is unusual to put water in a swimming pool, but because it would be unusual *not* to. Presumably the swimming pool frame constructed to understand (15) has a slot Contains(x) which has been filled by a default value Water. Interpretation of (15) indeed relies on knowledge of the typical backyard swimming pool. The first clause evokes a model of people having fun diving from a tree into a backyard swimming pool. Moreover, the second clause is initially interpreted as installing a piece of equipment commonly found near backyard swimming pools that might function in an analogous way to the tree. However, the disjunctor "water" prompts the reader to revise a default assumption of the Backyard Swimming frame, namely that there was water in the pool. Revising this simple assumption has substantial implications for the consequences of diving from the tree into the pool, and for the mindset of those who enjoy such activities. However, it is unlikely that these implications are represented in generic frames for Backyard Swimming Pools, and less likely that these implications can be logically derived from them.

Similarly, in (16) the reader's initial interpretation of the scenario relies on constructing a frame based on some combination of information provided by the text of the joke and background knowledge retrieved from long-term memory.

(16) By the time Mary had had her fourteenth child, she'd finally run out of names to call her husband.

Because parents typically name their children soon after they are born, the reader initially interprets "names" as baby-names. However, as "husband" is consistent with neither the listener's lexical level expectations, nor the message-level expectations which support them, it triggers frame-shifting. The frame-shifted interpretation, of course, involves reinterpreting "names" as derogatory epithets directed at the man Mary blames for her 14 pregnancies. Thus the lexical reanalysis of "names" triggers pragmatic renalysis of the frame constructed to support the initial interpretation.

In (16), both the invited and the joke interpretations concern the implication of having an atypical number of children, but while the invited reading concerns its implication for naming the children, the joke reading concerns its implication for the mother's quality of life. Moreover, *both* interpretations rely on knowledge of the events surrounding childbirth, just slightly different information. In this example, the frame-shifted interpretation requires knowledge that giving birth is painful, and that husbands play a causal role in their wive's pregnancies. The reanalysis of (16) thus demonstrates the way context can affect the interpretation of a word by calling to mind particular aspects of its conceptual structure (as in Barclay, Bransford, Franks, McCarrell, & Nitsch, 1974; Johnson-Laird, 1993; Tabossi, 1983; Tabossi & Johnson-Laird, 1980).

Further, while both analyses appeal to stereotypical knowledge of the sort frames typically represent, the scenario described in (16) is far from typical. First, it is quite atypical to have 14 children. Second, swearing at the husband is presumably not part of the scripts for Pregnancy and/or Childbirth. Moreover, swearing suggests a negative attitude towards childbearing, which – however warranted it may be – is not part of the scripted response to these events. On the one hand, (16) suggests that in the absence of information to the contrary, readers assume that typical assumptions obtain. However, when given information to the contrary, they are nonetheless able to adapt their models in order to incorporate new information. While a traditional script- or frame-based system can generate a new slot in response to an unexpected event, it is unable to compute the relationship between unexpected and normal events, because its inferencing capacity is based in knowledge represented in the restaurant script itself.

Given that one of the main difficulties of previous frame-based systems is in defining criteria for frame-selection (see Allen, 1987), the psychological reality of frame-shifting would seem to pose a serious problem for a frame-based account. In many cases, there is no previously stored frame which can be recruited to relate events to one another. The challenge of frame-shifting is creating a new superframe and adapting previously created structure accordingly. Given the rigidity of the frame as a data structure (e.g. Allen, 1987; Wilensky, 1986), one might question whether frame-based models can accommodate the demands of frame-shifting.

6.2 Connectionist Frames

The need for a sufficiently flexible implementation of frames has driven some researchers to explore the adequacy of subsymbolic processing in neural networks (see McClelland & Rumelhart, 1986 for review). Rumelhart, Smolensky, McClelland, & Hinton (1986) argue that frames are not static knowledge structures, but stable states in the energy landscape of constraint satisfaction networks. Units in the network represent semantic microfeatures, and the weights between units encode correlations between those microfeatures. The network is set up to promote excitatory weights between microfeatures which cooccur, and inhibitory weights between features which do not. Consequently, when one of the units is clamped on by the programmers, the network (using a gradient descent algorithm) fills in the frame by activating correlated microfeatures and inhibiting the others until it settles into a stable state.

In such a system, the rigidity of a frame will depend on correlations between the activation of its constituent microfeatures. If the network has experienced a high correlation between the mutual activation of stove, refrigerator, and kitchen, it will result in a narrow energy landscape that behaves like a rigid frame. On the other hand, if the concomitant activation of these features displays more of a family resemblance pattern, the energy landscape will be broader, reflecting the increased set of inputs that can instantiate the frame. The main advantage of connectionist implementations of frames over traditional symbolic ones is flexibility in accommodating inputs that do not conform to previously instantiated schemas. For example, in Rumelhart et al.'s (1986) model, even central elements can be missing and the network will nonetheless settle into a stable state. However, the particular network described by Rumelhart et al. would fail miserably in the representation of information in the sentences and jokes discussed above because it contains no mechanisms for generating the high-level inferences that relate frames to one another.

An example of a connectionist model which *is* able to draw high level inferences is St. John's (1992) Story Gestalt model which combines information from multiple sources to fill in aspects of the message-level (or "gestalt") representation. For example, given a story about dinner in a restaurant, the model can combine information about who paid the bill, whether that character is cheap or extravagant, and the quality of the restaurant, in order to predict the size of the tip. Rather than the 2-layer perceptron used by Rumelhart et al., St. John uses a single recurrent network (as in Elman, 1990) whose architecture is more suited to time series predictions in script-based reasoning. However, much like Rumelhart et al.'s (1986) model, the Story Gestalt model relies on cooccurrence frequencies in its input to infer default information.

For instance, consider the Bar script:

[Person1] decided to go to the bar.
[Person1] made a polite/obnoxious pass at [person2].
[Person2] gave kiss/slap to [person1].
[Person1] rubbed lipstick/cheek.

Given the information in the second event in this script, the model should be able to predict the probability of the last action as a 70:30 chance of a rubbed cheek (after an obnoxious pass), and a 30:70 chance of rubbed lipstick (after a polite pass). With the knowledge of the third event in the script, the model should be able to predict the last event deterministically with perfect validity. Moreover, when the model encounters an unexpected event, such as a kiss after an obnoxious pass, it is quite able to modify its prediction from “Person1 rubs cheek” to “Person1 rubs lipstick.”

St. John (1992) thus provides a model that illustrates how reanalysis of local predictive inferences can be constrained by higher-level information in a script. However, this reanalysis is not an example of true frame-shifting as only one higher-level frame is being recruited. Moreover, in spite of the flexibility afforded by connectionist-style modelling, the frame-shifting ability of the Story Gestalt mode is limited in much the same way as symbolic implementations. Because it is unable to compute the relationships between different higher-level representations, St. John’s (1992) model is incapable of combining information from different scripts in any sensible way. While distributed networks such as those described by Rumelhart et al. (1986) and St. John (1992) are well-suited for learning categories or other functions which can be induced from cooccurrence statistics, they often lack the representation of structure needed to understand conceptual relationships between elements in a discourse (Feldman, 1989).

6.2.1 Structured Connectionism

In an attempt to combine the representational capacities of symbolic NLP systems with the processing advantages of parallelism, Lange & Dyer (1989) argue for an alternative framework known as *structured connectionism.* In a structured network, each node stands for a distinct concept, and connections between nodes represent structural relationships between concepts. Lange & Dyer’s (1989) model ROBIN uses connections between nodes to encode semantic knowledge represented in a frame type data structure. Each frame has one or more slots, and slots have constraints on the type of fillers to which they can be bound. The relationships between frames are represented by excitatory and inhibitory connections between nodes and pathways between corresponding slots. Once initial role assignments have been made, ROBIN propogates evidential activation values in order to compute inferences from the information the programmers have given it.

Inference occurs as spreading activation propagates evidential activation across the excitatory and inhibitory connections which exist between related frames and competing slot-fillers. For example, connections between frames for Transfer-Inside and Inside-of allow the system to ‘infer’ Inside-of(Oven, Pizza) from Transfer-Inside(Pizza, Seana, Oven). In this model, frame selection is entirely a matter of spreading activation. Because each slot has a number of binding nodes, all of the meanings of an ambiguous word can serve as candidate bindings. Candidate bindings can be simultaneously propagated and the binding node with the greatest evidential activation eventually wins out. Because multiple frames are activated in parallel, contextual information can further activate an already highly activated node (or set of nodes), thus confirming an initial interpretation. Alternatively, contextual information can activate a previously less-active interpretation, thus implementing frame-shifting.

While the use of spreading activation to implement frame-shifting looks promising (as does the related model presented by Shastri & Ajjangadde, 1993), Lange & Dyer’s model has some serious deficiencies. First, ROBIN can’t represent scenarios that require recursion.

Second, it has no mechanism for adding new concepts, other than having a programmer add in new frames and their connections to the model's other frames. Third, ROBIN is unable to handle metonymic or metaphoric uses of language. Nor is it able to adapt frames to accept new fillers as in the concept ("same sex marriage" c.f. Turner & Fauconnier, 1995). Moreover, while ROBIN can cope with lexical ambiguity, it is completely unable to deal with what Clark (1983) has called semantic indeterminacy. While the need for frame-shifting on Lange & Dyer's model is prompted by lexical and conceptual ambiguity, no such ambiguity exists in the case of (16). That is, we would not want to say that a scenario in which people have fun diving into a swimming pool is ambiguous between cases where the pool *does* have water, and cases where it does *not*. Finally, the representation of each concept with a single node limits the model's ability to represent the meaning shading found in natural language.

6.3 An alternative approach

In an alternative approach to the frame-selection problem, Schank (1982) introduced the idea of MOPs, or memory organization packets. MOPs are generalized clusters of *scenes*, the high-level components of scripts. For example, a MOP for **entering** might consist of a set of scenes for entering a grocery store, a dentist's office, and a restaurant. Rather than employing encapsulated scripts such as the restaurant script, Schank (1982) uses a combination of general MOPs which deal with exchanges and services, and more specific MOPs with knowledge about particular restaurants. Schank (1982) found that encoding information at various levels of specificity maintained the representational advantages of scripts, while allowing for more combinations and recombinations of scenes at various levels. Further, Kellerman, Broatzmann, Lim & Kitao (1989) have suggested that MOPs are more psychologically plausible memory representations, citing data on memory confusions between stories that recruit distinct, though similar, scripts.

We propose a generalization of Schank's (1982) proposal in which the representations underlying language – even at the most rudimentary levels – have important properties of frames such as slot/filler organization, default values, and hierarchical structuring. In our approach, grammar does not algorithmically specify context invariant meanings, but rather provides clues to help the language user construct a message-level representation (Fauconnier, 1997; Gernsbacher, 1990; Turner, 1991). Frames, in their most generic forms, can be viewed as the basis for grammatical constructions (Goldberg, 1995; Langacker, 1987). In formalisms such as cognitive (Langacker, 1987) and construction (Kay, 1997) grammar, grammatical phenomena reflect the operation of very abstract frames for construing the conceptual content evoked by linguistic utterances.

Moreover, Langacker (1996, ms) has suggested that abstract frames evoked by grammatical information can help the language user to assemble the message-level representation. More specifically, he shows how a number of different grammatical phenomena can be described as directing meaning construction by shifting the hearer's focus of attention within some context (a process known as *profiling*). While Langacker's ideas do not solve the frame selection problem described above, they do suggest how different patterns of profiling enable the language user to coordinate abstract grammatical frames with other structure, including specific frames evoked by particular words, perceptual information, and the existing message-level representation.

In interpreting (16), for example, the construction "By the time X, Y" invites the listener to create a frame which imputes shared causality in the occurrence of state X and state Y. In (16) the relationship between X (Mary having her 14th child) and Y (running out of names) initially involves the naming slot in the childbirth script. However, "husband" suggests that the instantiation of the Naming slot in the childbirth script is erroneous. Presumably because "names" has been profiled as the head of the NP modified by "to call her husband," the incongruous item first prompts lexical reanalysis of "names" which in turn leads the speaker to a new frame to relate X and Y. Information at various levels of abstraction helps to constrain the initial analysis and subsequent reanalysis.

More specifically, interpretation of (16) involves recruiting and coordinating frames at various levels of abstraction, including the schematic relationship set up in the "By the time X, Y" construction, knowledge of events surrounding childbirth, and novel frames for the relationship between Mary, her husband, and her children. Example (16) thus demonstrates how context can affect lexical interpretation by calling up particular facets of background knowledge about the domain in question, and the need to go beyond generic information represented in frames as traditionally construed.

6.4 Summary

We have argued that linguistic utterances cue the retrieval of abstract grammatical frames, which speakers unify with frames evoked by lexical and contextual information. Meaning construction thus consists of constructing a series of simple cognitive models, while keeping track of common elements and relations in successive models. We examined the use of frames in the context of reading jokes and other sentences which require knowledge-based semantic reanalysis. Frame-shifting seems to occur when it is necessary to represent the relationship between two or more objects, actions, or events. If the disjunctor, or frame-shifting trigger, cannot be incorporated into existing structure, the words that served to evoke that structure are reanalyzed to provide a coherent bridge between the initial and the revised representations. The relationship between the disjunctor and the connector can be suggested by grammatical clues, conceptual relationships, or a combination of the two. The role of background information in such a system goes way beyond that of filling in gaps: integrating current experience with background knowledge is the very essence of comprehension.

The presence of an Ending Type effect in all three experiments, where people spent longer reading Joke than Nonjoke Endings, as well as the observation tat people spent loner on Shift than Nonshift Endings, all point to the psychological reality of the frame-shifting process. The observed results are thus consistent with the frame-based approach to comprehension sketched in the introduction. On this framework, understanding natural language utterances involves the construction of cognitive models in working memory by recruiting frames from long-term memory. In view of the computational complexity of frame-shifting, perhaps the most remarkable finding of these experiments is not the existence of an effect of frame-shifting, but that the effect was only on the order of a few hundred milliseconds. Given the dramatic reorganization of the conceptual representation necessary to understand even simple jokes, this relatively small difference in reading times is quite remarkable.

References

[1] Allen, J. (1987). *Natural Language Processing.* Philadelphia: Benjamin/ Cummings Publication.

[2] Attardo, S., D.H. Attardo, P. Baltes, and M.J. Petray. (1994). The linear organization of jokes: Analysis of two thousand texts. *Humor: International Journal of Humor Research* 7(1): 27-54.

[3] Barclay, J., J.D. Bransford, J.J. Franks, N.S. McCarrel, and K.E. Nitsch. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior* 13(4): 471-481.

[4] Barsalou, L.W. (1992). Frames, concepts, and conceptual fields. In Adrienne Lehrer and Eva Feder Kittay (eds.), *Frames, Fields, and Contrasts: New Essays in Lexical and Semantic Organization.* Hillsdale, NJ: Erlbaum.

[5] Bloom, P.A. and I. Fischler. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition* 8: 631-642.

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

[7] Clark, H.H. (1983). Making sense of nonce sense. In Giovanni B. Flores D'Arcais and R.J. Jarvella (eds.), *The Process of Language Understanding.* Chichester: John Wiley & Sons, 297-332.

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

[9] Fauconnier, G. (1997). *Mappings in Thought and Language.* Cambridge: Cambridge University Press.

[10] Feldman, J.A. (1989). Neural representation of conceptual knowledge. In L. Nadel, L.A. Cooper, P. Cullicover, and R.M. Harnish (eds.), *Neural Connections, Mental Computation.* Cambridge, MA: MIT Press, 68-103.

[11] Fillmore, C.J. (1968). The case for case. In E. Bach and R.T. Harm (eds.), *Universals of Linguistic Theory*, New York: Holt, Rinehart and Winston, 1-90.

[12] Fillmore, C. (1982). Frame semantics. In Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm.* Seoul: Hanshin. 111-137.

[13] Frazier, L. (1995). Issues in representation in psycholinguistics. In Joanne Miller and Peter Eimas, (eds.), *Speech, Language, and Communication: Handbook of Perception and Cognition*, 2nd edition. San Diego: Academic Press, 1-27.

[14] Gernsbacher, M.A. (1990). *Language Comprehension as Structure Building* Hillsdale, NJ: Lawrence Erlbaum Associates.

[15] Goldberg, A.E. (1995). *Constructions: A construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

[16] Graesser, A.C., Woll, S.B., Kowalski, D.J., & Smith, D.A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning and Memory* 6: 503-515.

[17] Johnson-Laird, P.N. (1993). The mental representation of the meaning of words. In A.I. Goldman (ed.), *Readings in Philosophy and Cognitive Science*. Cambridge, MA and London: MIT Press. 561-584.

[18] Just, M.A., P. Carpenter, and J.D. Wooley. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111(2): 228-238.

[19] Kay, Paul. (1990). Even. *Linguistics and Philosophy* 13(1), 59-111.

[20] Kellerman, K., S. Broatzman, T-S. Lim, and K. Kitao. (1989). The conversation MOP: Scenes in the stream of discourse. *Discourse Processes* 12: 27-61.

[21] Langacker, R.W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.

[22] Langacker, R. (1996). A dynamic account of grammatical function. Unpublished manuscript.

[23] Lange, T.E. and Dyer, M.G. (1989). High-level inferencing in a connectionist network. *Connection Science* 1: 181-217.

[24] McClelland, J.L. and D.E. Rumelhart, (eds.). (1986). *Explorations in parallel distributed processing: A Handbook of Models, Programs, and Exercises*. Cambridge, MA: MIT Press.

[25] McKoon, G. and Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology* 49, 25-42.

[26] Minsky, M. (1975). Frame system theory. In P.N. Johnson-Laird and P.C. Wason (eds.), *Thinking: Readings in Cognitive Science*. Cambridge: Cambridge University Press. 355-376.

[27] Minsky, M. (1980). Jokes and the logic of the cognitive unconscious. AI Memo No. 603. Massachusetts Institute of Technology.

[28] Pickering, M. and M. Traxler (1995). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35(3): 454-475.

[29] Rumelhart, D.E. (1980;1987). Understanding understanding. Paper presented at international Workshop on "Psycholinguistic Models on Production", held in Kassel, Federal Republic of Germany, from July 13-17. In Hans W. Dechert and Manfred Raupach (eds.), *Psycholinguistic Models of Production*, 209-228.

[30] Rumelhart, D.E., and A. Ortony. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro and W.E. Montague (eds.), *Schooling and the Acquisition of Knowledge*. Hillsdale, NJ: Erlbaum.

[31] Rumelhart, D.E., P. Smolensy, J.L. McClelland, and G. Hinton. (1986). Schemata and sequential thought processes in PDP models. In D.E. Rumelhart and J.L McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2: *Psychological and Biological Models*. Cambridge, MA: MIT Press, 7-57.

[32] Sanford, A.J. and S.C. Garrod. (1980). A demonstration of the situational basis of text-comprehension through implicit assignments of roles to entities. Unpublished manuscript, Dept. of Psychology, University of Glasgow.

[33] Sanford, A.J. and S.C. Garrod. (1981). *Understanding Written Language: Explorations Beyond the Sentence*. Chichester: John Wiley & Sons.

[34] Schank, R.C. and R.P. Abelson. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[35] Schank, R.C. (1982). Reminding and memory organization: An introduction to MOPS. In W.G. Lehnert and M.H. Ringle (eds.), *Strategies for Natural Language Processing*. Hillsdale, NJ: Lawrence Erlbaum.

[36] Schwanenflugel, P.J. and E.J. Shoben. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language* 24: 232-252.

[37] Shastri, L. and Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings using temporal synchrony. *Behavioral & Brain Sciences* 16(3):417-494.

[38] Sharkey, N.E. and Mitchell, D.C. (1988). Word recognition in a functional context: The use of scripts in reading. *Journal of Memory and Language* 24: 253-270.

[39] Simpson, G.B. (1994). Context and the processing of ambiguous words. In Morton Ann Gernsbascher (ed.), *Handbook of Psycholinguistics*. San Diego: Academic Press, 359-374.

[40] Spivey-Knowlton, M. and M. Tannenhaus. (1994). Referential context and syntactic ambiguity resolution. In Charles Clifton, Jr. and Lynn Frazier, and Keith Rayner (eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ: Lawrence Erlbaum.

[41] St. John, M.F. (1992). The story gestalt: A model of knowledge intensive processes in text comprehension. *Cognitive Science* 16: 271-306.

[42] Tabossi, P. and P.N. Johnson-Laird. (1980). Linguistic context and the priming of semantic information. *Quarterly Journal of Experimental Psychology* 32: 595-603.

[43] Tabossi, P. (1983). *Interpreting words in context.* D. Phil. Thesis, University of Sussex.

[44] Travis, C. (1981). *The True and the False: The Domain of the Pragmatic.* Amsterdam: Benjamins.

[45] Turner, M. (1991). *Reading Minds: The study of English in the age of Cognitive Science.* Princeton, NJ: Princeton University Press.

[46] Turner, M. and G. Fauconnier. (1995). Conceptual integration and formal expression. *Metaphor and Symbolic Activity* 10(3): 183-204.

[47] Wilensky, R. (1986). Points: a theory of the structure of stories in memory. In B.J. Grosz, K.S. Jones, and B.L. Webber (eds.) *Readings in Natural Language Processing.* Los Altos, CA: Morgan Kaufman Publishing, 459-473.

[48] Zwaan, R.A. and Radvansky, G.A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin* 123: 162-185.