Consider how one might understand the following ad, “For Sale, Parachute. Used once; never opened; small stain.” Although there are no explicit cues to do so, the reader naturally recruits knowledge about how and why parachutes are normally used in order to construe the listed features of this chute as being causally related to one another. Empirical support for the importance of such causal and relational information in language comprehension comes from the consistent finding that memory for short texts includes inferred material (Sanford 1981). Further, computational considerations imply that language comprehension involves a capacity for dynamic inferencing based on general knowledge represented as frames (Lange 1989).

Invented somewhat independently by researchers in the field of linguistics (Fillmore 1968), psycholinguistics (Sanford 1981), artificial intelligence (Minsky 1975), and natural language processing (Schank 1977), frame-type data structures are often invoked to account for inferential aspects of language comprehension (Fillmore 1982). As used here, frames are representations with slot/filler structure, default values, and weak constraints on the type of fillers for a given slot (Barsalou 1992). 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.

In linguistics, frame semantics is a research program in which a word’s semantic properties are described with respect to the way that they highlight aspects of an associated frame. For example, buy and sell both evoke the Commercial Transaction frame. But buy highlights the buyer and the goods, while sell highlights the seller and the money. Background knowledge thus figures prominently in the establishment of meaning, as language functions against the backdrop of conceptual structure. Below we examine the utility of frame semantics for models of language comprehension, and focus on its capacity to explain the sorts of inference needed to understand the parachute joke.

In section 1, we consider the role that pragmatics plays in on-line meaning construction, contrasting the view that context-independent meaning is supplemented with background and contextual knowledge with the view that background and contextual knowledge is essential for the construction of meaning.
Although most people working in pragmatics subscribe to varying versions of the latter, these portraits of the semantics/pragmatics interface have had almost no influence on language researchers in psychology, computer science, and cognitive science. One exception to this trend is the space structuring model (Coulson 2001), a model of language comprehension inspired by ideas in cognitive linguistics. Section 2 outlines some assumptions of the model, and sections 3 and 4 describe experimental work on joke comprehension designed to test some of those assumptions. Section 5 reviews the adequacy of some frame-based models in cognitive science, and offers some speculations about the future of this approach.

1 Pragmatics: What’s that?

On the classical approach to language, pragmatics is defined largely by what semantics leaves out (Gazdar 1979). This definition involves positing an implicit model of comprehension in which purely linguistic knowledge of meaning is used to determine sentence meaning. To determine utterance meaning, all that remains is to bring background knowledge and contextual information to bear on sentence meaning. Indeed the core phenomena in pragmatics are deixis, speech acts, and implicature – all topics in which the primary concern is how contextual factors modulate interpretation.

One problematic aspect of the classical view of meaning is that it relegates pragmatics to the role of disambiguator. Katz and Fodor (1963), for example, argued that a viable theory of semantics would require a theory of pragmatics that would disambiguate sentences vis a vis their contexts of utterance. In a sentence about a pawn in a chess game, pragmatics is supposed to answer the question “Which pawn?” Similarly, a prominent function of pragmatics on the traditional view is to give content to indexicals. Thus, on such an account, pragmatics is what answers questions such as “Who are you?” and “Where is here?”

Although few people working in pragmatics are likely to agree to exactly this characterization, unfortunately, it still captures the way that many linguists and psycholinguists think about language in context. In psycholinguistics, in particular, context is understood almost exclusively in terms of its role in the resolution of ambiguities in lexical meaning (Duffy et al. 2001) and syntactic structure (Clifton and Staub 2008; Frazier 1995; Spivey-Knowlton 1994). To the psycholinguist, then, pragmatics is what answers the question, “Is that funny ‘haha’ or funny ‘weird’?” Similarly, when confronted with the statement “Tonight
we’ll discuss sex with the President,” pragmatics is what tells us whether the
President will be the topic of discussion, or a participant in it.

While these are all valuable contributions, and disambiguator is a perfectly
respectable occupation, it is not (I believe) the role that background and con-
textual knowledge in fact plays in meaning construction. Though presumably
motivated by the observation that many words have many meanings, the prag-
matics as disambiguator model is ultimately undermined by careful consideration
of the way in which word meaning changes from context to context. Although
tomato is not ambiguous in the examples below, Johnson-Laird has pointed out
that different features of tomatoes are salient in (1) than in (2) (Johnson-Laird
1993).

(1) The tomato rolled across the floor.

(2) He accidentally sat on a tomato.

Examples such as this would seem to argue against a view in which the speaker’s
knowledge of the context of utterance helps her to adjudicate between a finite
set of determinate meanings. Rather, background knowledge about tomatoes
is used to enrich the cognitive models constructed to represent the referential
situation depicted in the sentences. Moreover, various pragmatic functions are
available that allow any noun to refer to an infinite set of related phenomena.
For example, BMW can be used to refer to all sorts of BMW-related things, in-
cluding stock in the company, as in (3), the building that houses the company,
as in (4), or shattered BMW car parts as in (5).

(3) He bought 6,000 shares of BMW.

(4) He works next door at BMW.

(5) My ex-planted a bomb in my car and I’ve been cleaning up BMW ever
since.

Similarly, Clark has pointed out that these sorts of mechanisms allow speakers
to improvise word meanings ad infinitum (Clark 1983). For example, in an article
about choosing a college roommate based on the appliances she owns, “com-
puter” can mean roommate who owns a computer, as in (6).

(6) Our computer fell in love and left school.
The problem with the pragmatics as disambiguator model is that while it presumes that computing utterance meaning is difficult because language is semantically ambiguous, the difficulty really lies in the fact that language is semantically indeterminate. Rather than helping to adjudicate between pre-specified meanings, it would seem that context plays a constitutive role in the construction of meaning. Moreover, any mechanism that can account for the way that people can deal with semantic indeterminacy, ought to account quite trivially for semantic ambiguity as well. Rather than attempting to explain how pragmatic factors disambiguate a fully specified meaning constructed by semantics, then, it makes more sense to explore how pragmatic mechanisms contribute to the specification of meaning in the first place.

2 Space structuring model

On the space structuring model (Coulson 2001), a model of language comprehension motivated by ideas in mental space theory (Fauconnier 1994), conceptual blending theory (Fauconnier and Turner 1998; Fauconnier and Turner 2002), and cognitive grammar (Langacker 1987), perceptual input, language input, social context, and the speaker's current cognitive state all contribute to the construction of cognitive models of the discourse situation. This can include models of the referential aspects of sentences as well as models relevant to the agent's social and material goals. Both linguistic and non-linguistic cues prompt the retrieval of frames from long-term memory, and these frames are exploited in the construction of cognitive models of the message-level representation.

Three assumptions of the model include:

(i) the embodiment assumption, that the structure of language at least partially reflects bodily constraints on perception and action;
(ii) the immediacy assumption, that the integration of linguistic and non-linguistic information occurs rapidly, and does not (necessarily) require the prior construction of a propositional representation of sentence meaning;
(iii) the elaboration assumption, that language comprehension involves animating the cognitive models constructed by the listener.

In traditional frame-based approaches (Schank 1977), comprehension requires frames to be bound to contextually available elements. However, in the space structuring model, this need not be the case. At times comprehension can proceed by binding slots or attributes in the activated frame. However, often the frame serves only to constrain the construction of a cognitive model that is
particularized to the discourse situation. Like the frames that inform them, these
cognitive models are hierarchically organized, have attribute/value structure,
and a mechanism that assigns default values for unspecified attributes. Though
schematic and partial, these models are detailed enough to enable small-scale
simulations of the scenarios they represent (as in the mental models described
by Norman 1974).

For example, if a listener heard the sentence in (7), she might, at one level
at least, respond by constructing a model of the referential situation described
by the speaker.

(7) When I asked the bartender for something cold and full of rum, he recom-
mended his daquiri.

Of course, at other levels the listener might be building models related to why
the speaker might make this particular statement, what the speaker’s attitude
is toward her, or many other things. Nonetheless, at the referential level, the
listener combines linguistic information with background knowledge to build a
cognitive model of an interaction between a customer and a bartender.

The fact that speakers build models like this, and the extent to which those
developing models guide their expectations is perhaps best appreciated by
examining examples in which those expectations are violated. For example, (8)
is very similar to (7) but prompts the construction of a very different cognitive
model.

(8) When I asked the bartender for something cold and full of rum, he recom-
mended his wife.

Rather than recommending a drink, the bartender in (8) has just insulted his
wife. The semantic and pragmatic reanalysis that reorganizes existing elements
in the message-level representation is known as frame-shifting (Coulson 2001).
With the activation of background knowledge and the establishment of mappings
between counterpart structure in the old frame and in the new one, the bar-
tender’s wife is accused of being a frigid lush.

In fact, jokes are deliberately constructed to suggest one frame while evok-
ing 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. For example, in (9) the reader begins by
evoking a frame in which a busy professional pays an accountant to do his
taxes.
(9) I let my accountant do my taxes because it saves time: last spring it saved me ten years.

The disjunctive *years*, however, forces the reader 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 (9) requires recruitment of background knowledge about the particular sorts of relationships that can obtain between business people and their accountants so that the initial busy professional interpretation can be mapped into the crooked-businessman frame.

Frame-shifting involves a dramatic reorganization of the message-level aspects of the utterance, most of which can’t be attributed to compositional mechanisms of reanalysis. For example, most examples of frame-shifts in jokes don’t require the listener to instantiate a new structural analysis of the sentence. Though the listener is led down a pragmatic garden path, it is often the case that pragmatic reanalysis proceeds without syntactic reanalysis. In (8), for example, *wife* is the object of *recommended* just as *daquiri* is in (7), the straight version of (8). In (9), the joke interpretation actually requires the reader to abandon the fully grammatical reading of “saved me ten years” for something akin to “saved me jail time,” which is questionable at best.

Further, such cases of frame-shifting frequently require the creation of nonce senses. For example, in the accountant joke in (9) “saves time” is re-interpreted as meaning “prevents me from having to do time.” In the case of the bartender’s wife, “full of rum” comes to mean “alcoholic.” However, the construction of these somewhat novel phrasal meanings is as much the effect of frame-shifting as the cause. That is, it seems likely that understanding the novel reading of “full of rum” at least partially depends on the construal of the bartender’s speech act as an insult. Moreover, the adaptation of the idiomatic meaning of *time* in (9) as in (“do time”) is only congruous because of a stereotyped scenario that involves accountants obscuring illegal business dealings.

### 3 The psychological reality of frame-shifting

Demonstration of the psychological reality of frame-shifting, then, would suggest a role for pragmatics that goes beyond pragmatics the disambiguator. In
particular, frame-shifting suggests that lexical processing does not simply benefit from context, but actively contributes to it. Moreover, background and contextual knowledge do not merely help the listener to specify the meaning of indexicals and disambiguate the meaning of lexical items, but, rather, are crucial for the construction of the message-level representation. Below we discuss the results of studies using three different techniques that establish the psychological reality of frame-shifting: self-paced reading times, eye movement registration, and event-related brain potentials.

3.1 Self-paced reading times

To demonstrate the psychological reality of frame-shifting, Coulson and Kutas (1998) conducted a variety of experiments using the self-paced reading time technique. In this experimental paradigm, the task is to read sentences one word at a time, pressing a button to advance to the next word. As each word appears, the preceding word disappears, so that the experimenter gets a record of how long the participant spent reading each word in the sentence.

Stimuli for this experiment were comprised of one-line jokes that required frame-shifting for their comprehension, and straight versions of the same sentences that did not require a frame-shift. Moreover, because we wanted to be able to detect the effect of frame-shifting on the processing of a single word, the disjunct, or frame-shifting trigger, was always a sentence-final noun. In order to find out what sort of non-joke frames people constructed for these sentences, we performed a norming task in which people were given the jokes minus the last word and asked to complete the sentence with the first word or phrase that came to mind. This is known as a cloze task, and the percentage of people who offer a given word in a given sentence context is known as the cloze probability of that word in that particular sentence context.

Results of the cloze task enabled Coulson and Kutas to ascertain readers’ default (non-joke) interpretation for the sentences. However, it also revealed a disparity in the cloze probability of the most popular response for the items, suggesting that some of the sentence fragments provided a more constraining context than others. For example, (10) elicited a similar response from 81% of the participants, while (11) elicited many different responses, albeit mostly from the gambling frame.

(10) I asked the woman at the party if she remembered me from last year and she said she never forgets a (face 81%).
(11) My husband took the money we were saving to buy a new car and blew it all at the (casino 18%).

As a result, two types of jokes were tested: high constraint jokes like (10) which elicited at least one response with a cloze probability of greater than 40%, and low constraint jokes like (11) which elicited responses with cloze probabilities of less than 40%. To control for the fact that the joke endings are (by definition) unexpected, the straight controls were chosen so that they matched the joke endings for cloze probability, but were consistent with the frame evoked by the context. For example, the straight ending for (10) was name (the joke ending was dress); while the straight ending for (11) was tables (the joke ending was movies). The cloze probability of all four ending types (high and low constraint joke and straight endings) was equal, and ranged from 0% to 5%.

Given the impact of frame-shifting on the interpretation of one-line jokes, one might expect the underlying processes to take time, and, consequently result in increased reading times for jokes that require frame-shifting than “straight" versions of the same sentences. Coulson and Kutas (1998) found that readers spent longer on the joke than the straight endings, and that this difference in reading times was larger and more robust in the high constraint sentences. This finding suggests there was a processing cost associated with frame-shifting reflected in increased reading times for the joke endings, especially in high constraint sentences that allow readers to commit to a particular interpretation of the sentence.

3.2 Eye movement registration

The self-paced reading paradigm is a good technique for establishing that one kind of sentence requires more processing time than another (presumably very comparable) type of sentence. However, one drawback to this technique is its lack of ecological validity. In contrast to normal reading, participants in a self-paced reading task are permitted to see only one word at a time, and moreover, are not permitted to look back at earlier regions of the sentence. In contrast, in free reading, people frequently move their eyes leftward (or regress) to re-examine earlier parts of the text. Another deficit of the self-paced reading paradigm is that it provides little information about the nature of the processing difficulty that readers encounter. For example, longer reading times for sentences that ended as jokes than straight controls in Coulson and Kutas (1998) suggests the jokes engendered more processing difficulty. However, reading time data do not indicate whether this difficulty occurs in the initial stages...
of word processing, or later as the reader moves on to inferential aspects of
processing.

To address these questions, Coulson, Urbach and Kutas (2006) conducted an
eye movement study comparing reading times for sentences that ended as jokes
to reading times for the same sentences with the unexpected straight endings
(Coulson and Kutas 1998). They found that just as in the reading time study,
readers spent reliably longer on the joke endings, and that this difference was
far more pronounced for the high constraint sentences. Moreover, this effect
arose in the later stages of processing associated with subsequent fixations of
the sentence final word (rather than in the initial phase of processing). Further,
in both high and low constraint sentences, participants were more likely to
regress when they encountered joke endings than straight ones. This finding
is consistent with the psychological reality of frame-shifting, suggesting readers
literally revisit aspects of the preceding context in order to get the jokes.

3.3 Event-related brain potentials

Another way of assessing readers’ on-line comprehension of language materials
is to use event-related brain potentials (henceforth ERPs). ERPs provide an
on-going record of brain activity related to various kinds of sensory, motor, and
cognitive processing events. The physical basis of the ERP signal is the fact that
when large groups of neurons (on the order of tens of thousands) fire simultane-
ously, they create an electrical field in the brain that can be detected with electro-
dodes at the scalp via the electroencephalogram (henceforth EEG). The ERP is
obtained by applying electrodes to the scalp, recording participants’ EEG, and
averaging across events within experimental categories. Because the averaging
process presumably cancels out the EEG that is not related to the experimenter’s
categories, the remaining signal represents the brain activity related to the
processing of the experimental stimuli. By comparing the ERPs to different sorts
of stimuli, the researcher can assess how changing the nature of the cognitive
task modulates the brain response.

Because eye movements necessary for normal reading produce artifacts in
the EEG, ERP reading experiments typically involve presenting sentences one
word at a time in the center of a computer monitor. EEG can thus be time-locked
to the onset of each word on the monitor, and the resultant ERP represents brain
activity associated with reading a particular category of words (i.e., the last word
of an incongruous sentence). The ERP is a waveform with a series of positive and
negative peaks (often called components) that can be correlated with various
types of processing. Components are generally labeled by reference to their
polarity (P for positive-going and N for negative-going activity), and their latency, or when they occur relative to either the onset of the stimulus event or to other ERP components.

In a classic ERP language experiment, Kutas and Hillyard (Kutas and Hillyard 1980) contrasted ERPs elicited by visually presented sentences that ended congruously, as in (12), with ERPs elicited by sentences that ended incongruously, as in (13).

(12) I take my coffee with cream and sugar.

(13) I take my coffee with cream and socks.

They found a negativity in the brainwaves that was much larger for incongruous sentence completions than the congruous ones. Because it peaks about 400 milliseconds after the onset of a visually presented word, this negativity is called N400.

Over forty years of research have revealed reliable relationships between the nature of various stimulus and task manipulations designed to alter participants’ cognitive state, and corresponding modulations of ERP components. For example, the P1 component is the first positive deflection in the ERP elicited by visually presented words. This component, evident 70–100 milliseconds after the word is shown (or post-word onset), reflects early sensory and vision-related attentional processing. Mangun, Hillyard and Luck (Mangun 1993) have proposed that the P1 component reflects a gating mechanism responsible for modulating the width of the attentional spotlight. P1, N1, and P2 components elicited during reading probably reflect the visual feature extraction necessary to relate the visual stimulus to information in memory (Kutas and King 1996).

Most ERP studies of language processing have focused on longer latency components, such as the N400. Because the N400 was initially reported as a brain response to incongruous sentence completions (Kutas and Hillyard 1980), many people mistakenly believe it is only elicited by semantic anomalies. However, research indicates it is a far more generally elicited ERP component associated with the integration of a word into the established context. In fact, N400 is elicited by all words, spoken, signed, or read, and its size, or amplitude, is an index of the difficulty of lexical integration. The best predictor of N400 amplitude is a word’s cloze probability in a particular sentence: N400 is small for high cloze expected completions like sugar in (12), large for low cloze completions like socks in (13), and intermediate amplitude for sentence final words of intermediate cloze probability.
Besides providing an on-going record of brain activity during language processing, ERP data can complement reaction time data such as that collected in the self-paced reading and eye movement registration paradigms discussed above. These two kinds of data are often complementary as reaction time data can provide an estimate of how long a given processing event took, while ERP data can suggest whether distinct processes were used in its generation. An experimental manipulation that produces a reaction time effect might produce two or more ERP effects, each of which is affected by different sorts of manipulations. To the extent that ERP effects can be identified with specific cognitive processes (i.e., the N400 and lexical integration), they provide some evidence of how processing differs in the different conditions (King 1995).

With this in mind, Coulson and Kutas (2001) recorded participants' brainwaves as they read sentences that ended either as jokes or with unexpected straight endings (Coulson and Kutas 1998). In an ERP study of the brain response to jokes like those discussed in the previous section, Coulson and Kutas found that ERPs to the joke endings differed in several respects from those to the straight endings, depending on contextual constraint as well as participants' ability to get the jokes. In poor joke comprehenders, jokes elicited a negativity in the ERPs between 300 and 700 milliseconds after the onset of the disjunctor. In good joke comprehenders, high but not low constraint endings elicited a larger N400 (300–500 ms post-onset) than the straights. Also, in this group, both sorts of jokes (high and low constraint) elicited a positivity in the ERP (500–900 ms post-onset) as well as a slow, sustained negativity over left frontal sites. Multiple ERP effects of frame-shifting suggest the processing difficulty associated with joke comprehension involves multiple neural generators operating with slightly different time-courses.

3.4 Summary

Taken together, these three studies of frame-shifting in jokes are far more informative than any one study alone. The self-paced reading time studies suggested that frame-shifting needed for joke comprehension exerts a processing cost that was especially evident in high constraint sentence contexts (Coulson 1998). The eye movement study of the same stimulus set confirmed that the processing cost of frame-shifting was evident under more natural reading conditions, and replicated the finding that differences between reading times for joke and straight endings were much larger for high constraint sentences. Moreover, the eye movement study suggested that the processing cost was not at the level of word recognition (indexed by the length of a reader's initial fixation of a word),
but was related to higher-level processing indexed by the total amount of time spent looking at the word (that is the sum of the time that elapsed during the initial fixation as well as all subsequent fixations). Coulson, Urbach and Kutas (2006) also found that people were more likely to make regressive eye movements when they read the joke than the straight endings, suggesting they wanted to re-examine earlier parts of the sentence for clues to which alternative frames should be retrieved.

ERP results from the study by Coulson and Kutas (2001) also suggest the processing cost associated with frame-shifting is related to higher-level processing. In the case of the high constraint jokes, the difficulty includes the lexical integration process indexed by the N400, as well as the processes indexed by the late-developing ERP effects. In the case of the low constraint jokes, the difficulty was confined to the processes indexed by the late-developing ERP effects. The added difference in lexical integration indexed by the N400 may explain why the joke effects on both reading times and gaze durations were more pronounced for high constraint sentences than for low. Because the late developing ERP effects were only evident for good joke comprehenders who successfully frame-shifted, they are more likely to be direct indices of the semantic and pragmatic reanalysis processes involved in joke comprehension. The temporally extended nature of these effects – lasting at least 400 ms – is also consistent with the idea that they index the construction of the message-level representation.

4 The neural substrate of frame-shifting

Another way of exploring the computational demands of frame-shifting is to look at the overlap in the brain regions underlying joke comprehension and those invoked for comparable language comprehension tasks. Interestingly, joke comprehension is thought to recruit brain regions above and beyond the set of left hemisphere (henceforth LH) areas thought to underlie core language abilities. Neuropsychologists have suggested that joke comprehension is particularly compromised in patients with right hemisphere (henceforth RH) lesions, especially when there is damage to the frontal lobe (Brownell et al. 1983; Shammi and Stuss 1999). Classic studies have assessed patients’ ability to choose the punch line for a joke from an array that includes both straightforward and non sequitur endings as distracter items. On such tests, right hemisphere damage (RHD) patients have tended to choose the non sequitur endings, suggesting they understood that jokes involve a surprise ending, but had
difficulty with the frame-shifting process required to re-establish coherence (Brownell et al. 1983).

The pattern of deficits in RHD patients differs dramatically from those evidenced by LHD patients whose communicative difficulties are seemingly more severe. To compare the performance of LHD and RHD patients on joke comprehension, Bihrl and colleagues used both verbal (jokes) and nonverbal (cartoons) materials with the same narrative structure (Bihrl et al. 1986). Patients were asked to pick the punch-line (or punch frame) from an array of four choices: a straightforward ending, a neutral non sequitur, a humorous non sequitur, or the correct punch-line. Though both LHD and RHD groups were impaired on this task, their errors were qualitatively different. RHD patients showed a consistent preference for humorous non sequitur endings over straightforward endings and correct punch-lines; in contrast, LHD patients more often chose the straightforward endings than either of the non sequitur endings (Bihrl et al. 1986). RHD patients displayed preserved appreciation of the slapstick depicted in the humorous non sequitur endings, but were impaired at the frame-shifting needed to understand the correct punch lines.

One attempt to link the deficits observed in RHD patients to hemispheric asymmetries evident in healthy adults is Beeman’s coarse coding hypothesis (Beeman and Chiarello 1998; Beeman et al. 1994). According to this hypothesis, words in the RH are represented by means of wide semantic fields, while words in the LH are represented via a narrow range of features relevant to the immediate discourse context. Because jokes frequently require the integration of novel information, the reinterpretation of a word or phrase, and the reinterpretation of the scenario depicted by the preceding context, diffuse RH activation might provide additional information that makes joke processing easier. Similarly, Coulson and Wu (2005) have suggested that semantic activations in the RH include causal and relational information crucial for frame-shifting. Reduced access to these semantic activations in RH damaged patients could result in joke comprehension deficits.

Several studies in our laboratory have addressed whether hemispheric differences in semantic activation are relevant for joke comprehension. In one study, we recorded ERPs as healthy adults read laterally presented “punch words” to one-line jokes (Coulson and Williams 2005). Lateral presentation of critical words was intended to affect which cerebral hemisphere received the initial information from the stimulus, and to increase the participation of that hemisphere in the processing of the stimulus. The organization of the visual system is such that words presented in the left visual field (henceforth LVF) are initially processed in the RH, while words presented in the right visual field (henceforth RVF) are initially processed in the LH. In healthy adults, the visual...
information is rapidly transferred to the other hemisphere. Nonetheless, lateral presentation shifts the balance of processing to favor the hemisphere that initially received the visual information, and as such can give us clues to processing differences between the two halves of the brain.

In our study of how lateral presentation impacts ERPs to jokes, the N400 component was of particular interest, as its amplitude indexes the difficulty of integrating the meaning of a given word into one’s model of the discourse context (Kutas and Hillyard 1980; Kutas and Van Petten 1994). As noted above, the critical word in a joke often elicits a larger N400 than a similarly unexpected “straight” ending for the same sentence: the N400 joke effect (Coulson and Kutas 2001). We reasoned that if hemispheric differences in semantic activation are relevant for joke comprehension, lateral presentation of joke (GIRL) versus straight (BALL) endings for sentences such as “A replacement player hit a home run with my” would result in different N400 joke effects as a function of visual field of presentation.

In this sentence comprehension paradigm, the difficulty of joke comprehension is indexed by the size of the N400 joke effect with larger effects pointing to relatively more processing difficulty. In fact, N400 joke effects were smaller when the critical words were presented to the LVF/RH than the RVF/LH, suggesting joke comprehension was easier with LVF presentation and consistent with the claim that coarse coding in the RH facilitates joke comprehension (Coulson and Williams 2005).

In a similarly motivated study, we measured ERPs elicited by laterally presented probe words that were preceded either by a joke, or by a non-funny control (Coulson and Wu 2005). Since all jokes turned on the last word of the sentence, control sentences were formed by replacing the sentence final word with a “straight” ending. For example, the straight ending for “Everyone had so much fun diving from the tree into the swimming pool, we decided to put in a little water,” was “platform.” Probes (such as crazy) were designed to be related to the meaning of the joke, but unrelated to the meaning of the straight control. In this sentence prime paradigm, the activation of information relevant to joke comprehension was signaled by differences in the size of the N400 elicited by related versus unrelated probes. The more active joke-related information was, the larger the N400 relatedness effect could be expected to be. Consistent with the coarse coding hypothesis, we found larger N400 relatedness effects with LVF/RH presentation suggesting joke-related information was more active in the RH (Hull, Chen, Vaid and Martinez 2005 for comparable evidence using behavioral measures).

The importance of the RH in understanding narrative jokes, however, contrasts with its importance in understanding the word play in puns. Coulson and
Severens (2007) addressed hemispheric sensitivity to the different meanings of a pun using a sentence prime paradigm with puns and pun-related probe words. We recorded ERPs as healthy adults listened to puns and read probe words presented in either participants’ left or right visual fields. Probe words were either highly related to the pun that preceded them, moderately related to the pun that preceded them, or were unrelated to the pun that preceded them. For example, the highly related probe for “During branding cowboys have sore calves,” was “cow” and the moderately related probe was “leg”.

The activation of pun-related information was assessed by the presence of relatedness effects on the N400 component of the ERP and on positive waveforms that frequently follow the N400 such as the late positive complex (henceforth LPC). With an ISI of 0ms, we observed similarly sized priming effects for both the highly and moderately related probes with RVF/LH presentation; with LVF/RH presentation, we observed priming for the highly but not the moderately related probes. With an ISI of 500 ms, we observed similarly sized N400 relatedness effects for highly and moderately related probes with presentation to the RVF/LH as well as the LVF/RH. In addition, RVF/LH, but not LVF/RH presentation, resulted in a larger centro-parietally distributed LPC for related probes. In sum, these results suggest that initially both meanings of a pun were equally active in the LH while only the highly related probes were active in the RH. By 500 ms after the offset of the pun, both meanings were available in both hemispheres.

The importance of the LH for pun comprehension thus contrasts with the role of the RH in understanding simple narrative jokes of comparable complexity in terms of vocabulary and grammar. Presumably this relates to the relative import of the retrieval of word meanings in puns versus frame semantic information involved in narrative jokes. While narrative jokes begin by suggesting one interpretation of the discourse situation only to replace it with another at the punch line (Giora 1991, 2003), the point of puns is simply to promote both meanings of an ambiguous word or phrase. The LH advantage observed in Coulson and Severens (2007) may reflect the importance of this hemisphere (especially the left frontal lobe) in coding the association between a word’s form and its meaning.

In fact, a neuroimaging study that compared narrative jokes with non-funny controls revealed bilateral temporal lobe activations, while an analogous comparison using puns revealed left frontal activations (Goel and Dolan 2001). Whereas the frontal activations to puns were consistent with the need to retrieve word meanings, the temporal lobe activations in both the left and the right hemispheres presumably reflect memory processes important for frame-shifting.
5 Integration

The research reviewed above suggests that the relationship between a word and its surrounding context is multifaceted. This relationship involves both the way that individual words add to the cognitive models active in working memory, and the way that individual words can prompt the construction of new models. Of course, the experiments reviewed above do not substantiate all of the theoretical claims we have made. They merely establish that people spend longer reading jokes than straight versions of the same sentences, that jokes prompt more regressive eye movements, that jokes elicit slightly different brainwaves, and that the recruitment of right hemisphere brain regions is more important for jokes that rely on frame-shifting, than for puns which do not. These findings are, however, consistent with the psychological reality of frame-shifting, and highlight the importance of this process for any comprehensive account of meaning construction.

Frame-shifting seems to occur whenever it is necessary to represent the relationship between two or more objects, actions, or events. If the disjunct or frame-shifting trigger, cannot be sensibly 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 disjunct and the connector can be suggested by grammatical clues, conceptual relationships, or a combination of the two.

Interestingly, frame-shifting presents a bit of a paradox for traditional frame-based models of language comprehension. On the one hand, the computational challenge of connecting an initial interpretation to the reinterpretation seems to require essential properties of frames, including the representation of causal and relational information, attribute/value (or slot/filler) organization, and the existence of default values. On the other hand, given the rigidity of the frame as a data structure (e.g. Allen 1987; Wilensky 1986), it is highly questionable as to whether frame-based models can accommodate the demands of frame-shifting.

5.1 Traditional implementations of frames

Wilensky (1986), for example, has argued that scripts are rigid data structures that cannot accommodate events that are out of the ordinary. While knowledge of typical scenarios represented in scripts and frames is necessary for understanding narrative jokes, it is far from sufficient. For example, in (14), the word
water is surprising, not because it is unusual to put water in a swimming pool, but because it would be unusual not to.

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

Presumably, the swimming pool frame constructed to understand (14) has a Contains(x) slot that has been filled by its default value water. Interpretation of (14) indeed relies on knowledge of the typical backyard swimming pool. The first clause (“Everyone had so much fun diving from the tree into the swimming pool,”) evokes a model of people having fun diving from a tree into a backyard swimming pool. Moreover, cloze data collected by Coulson and Kutas (1998) suggests the second clause (“we decided to put in a little . . .”) is initially interpreted as referring to the owner’s decision to install a piece of equipment commonly found near backyard swimming pools that might function in an analogous way to the tree (e.g. a diving board).

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.

The challenge of frame-shifting is to create a new super-frame and to adapt previously created structure accordingly. 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 on knowledge represented in the frame itself. In many cases, there is simply no frame that can be recruited to relate events to one another. Ironically, traditional implementations of frames are completely inadequate for modeling the semantic reorganization involved in frame-shifting.

5.2 Sub-symbolic implementations of frames

The need for a sufficiently flexible implementation of frames has driven some researchers to explore the adequacy of sub-symbolic processing in neural networks (McClelland 1986). The propensity of these networks to display (a) graceful degradation, viz. arriving at a best guess given imperfect information, (b)
spontaneous generalization, that is, accommodating inputs that do not conform
to previously instantiated schemas, and (c) the ability to arrive at a compromise
solution to mutual constraint satisfaction problems is compatible with the flexi-
bility people show in their interpretation of jokes.

The early promise of this approach can be seen in a model proposed by
Rumelhart and colleagues (Rumelhart 1986) that classifies rooms in a house
based on their contents (e.g. whether they have beds, chairs, refrigerators, and
so on). Units in the network represent semantic micro-features, and the weights
between units encode correlations between those micro-features. The network is
set up to promote excitatory weights between micro-features that co-occur, and
inhibitory weights between features that do not. If the network has experienced
a high correlation between the mutual activations of stove, refrigerator, and
counter, when the stove unit is activated, the network (using a gradient descent
algorithm) activates correlated micro-features (e.g., refrigerator, counter) until it
settles into a kitchen frame.

While the model by Rumelhart and colleagues reveals the flexibility of
probabilistic approaches, it is incapable of representing information needed to
get the jokes discussed above. This is because it contains no mechanisms for
generating the high-level inferences that relate frames to one another. A more
sophisticated model by St. John is able to use co-occurrence frequencies in its
input to infer default information, and to modify its predictions about upcoming
events in a way that is sensitive to context (St. John 1992). However, St. John’s
model is limited in much the same way as symbolic implementations: information
that deviates too much from stored frames cannot be accommodated.
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.

Lange and Dyer (1989) propose a structured connectionist model called
ROBIN (role-binding network) that explicitly attempts to capture inferential revi-
sions in frame-shifting. 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 the pathways between corre-
sponding slots. Once initial role assignments have been made, ROBIN propa-
gates evidential activation values in order to compute inferences from the infor-
mation the programmers have given it.

Inference is understood as resulting from the spread of activation across the
connections 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 (Pizza, Oven) from Transfer-Inside (Seana, Pizza, 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 activated simultaneously, 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.

The neurally inspired architecture of these models contributes important advances over traditional, symbolic implementations of frames. Advances include probabilistic representations, parallel activations, and the use of spreading activation mechanisms. However, none of these models have the capacity to creatively combine frames, to draw inferences that require an understanding of the relationship between frames, or to construct novel frames in response to contextual demands. While sub-symbolic implementations of frames represent an improvement over traditional frame-based models, they share many of the same limitations.

5.3 Grounded frames

The best way to achieve the flexible interpretative capacity needed for frame-shifting may be to adopt an empirically inspired architecture that is based on dynamic internal imagery (Bergen and Coulson 2006). In such models, language interpretation involves the creation of internal simulations of events that include sensory, motor, and affective dimensions. Barsalou (1999) has suggested that background knowledge is stored as perceptual symbols, schematic representations of perceptual experience stored around a common frame that promotes schematized simulations. Perceptual symbols are thus grounded in experience as the brain captures states across modalities and integrates them into a multi-modal representation stored in memory. This multi-modal representation is later reactivated to simulate relevant aspects of perception, action, and introspection.

Whereas frames have traditionally been understood as amodal knowledge structures that result from a distillation of experience, perceptual symbols are understood as having modal characteristics. Perceptual symbols recruit brain areas involved in the acquisition of the relevant concepts, and have some characteristics of analogue representations. However, perceptual symbols have also been argued to be schematic enough to implement standard symbolic functions, such as type-token relationships, recursion, and inference (Barsalou 1999).
As in traditional implementations of frames, perceptual symbols capture causal and relational information needed for frame-shifting. Moreover, because they are the product of neural learning mechanisms, perceptual symbols incorporate many of the features of sub-symbolic frames, such as partial and probabilistic activation patterns. Perhaps most importantly, the analogue character of perceptual symbols allows for novel combinations based on the affordances of the constituent concepts (Glenberg and Robertson 2000).

This model is supported by research indicating that the neural systems responsible for performing actions or perceiving percepts are also recruited for linguistically inspired simulations (Barsalou 2008). Consistent with the embodiment and elaboration assumptions in the space structuring model, recent findings suggest that language processing utilizes the perceptual and motor systems as internal models that allow for the construction of subjective experiences in the absence of motor action or perceptual input (Pecher and Zwaan 2005).

Bergen and Coulson (2006) argue that a simulation based model might account for the joke about the pool in (14). Because our experiences with pools almost without exception include water, water will automatically be activated in mental simulations that involve pools. Our experience with diving, by contrast, presumably involves some cases of landing on a solid surface, thus enabling us to viscerally imagine diving into a pool with water.

5.4 Conclusion

In accordance with the frame semantics program, 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 involves assembling a series of simple cognitive models while keeping track of common elements and relations. But, rather than just retrieving and instantiating frames, speakers are continuously and creatively building and blending cognitive models to yield new concepts and construals.

This view correctly predicts that scenarios which occasion frame-shifting present a challenge to the processor that differs from that presented by lexical violations consistent with the currently active frame. In contrast to the impoverished notion of context in psycholinguistics as something that is important only insofar as it facilitates processes that are clearly linguistic, the difficulty of frame-shifting in jokes demonstrates the need for a model of message-level processing prompted by language. Moreover, it suggests that message level representations are amenable to fairly substantive changes with minimal linguistic input.
Indeed the demands of joke comprehension suggest the models we build and revise so quickly derive from perceptual symbols, schematic representations of perceptual experience stored around a common frame that promotes schematized simulations (Barsalou 1999). We suggest that with frames built from perceptual symbols, one could maintain the representational advantages of hierarchically organized slot-filler structures, as well as explaining how speakers might construct a simulation of a parachute that was used, but never opened, in order to infer the origin of its stain.

References


Minsky, Marvin. 1980. Jokes and the logic of the cognitive unconscious (AI Memo No. 603). MIT.


